U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA)

+ + + + +

FEDERAL INSECTICIDE, FUNGICIDE AND RODENTICIDE ACT SCIENTIFIC ADVISORY PANEL (FIFRA SAP)

+ + + + +

REEVALUATION OF THE HUMAN HEALTH EFFECTS OF ATRAZINE: REVIEW OF EXPERIMENTAL ANIMAL AND IN VITRO STUDIES AND DRINKING WATER MONITORING FREQUENCY

+ + + + +

DOCKET NO.: EPA-HQ-OPP-2010-0125

+ + + + +

THURSDAY,

APRIL 29, 2010

+ + + + +

The Panel convened at 8:30 a.m. in the Hamilton Ballroom of the Hamilton Crowne Plaza

Hotel, located at 1001 14th Street, N.W., Washington, D.C., Steven G. Heeringa, Ph.D., Chair, and Kenneth M. Portier, Ph.D., Session Chair, presiding.

FIFRA SAP MEMBERS PRESENT:

STEVEN G. HEERINGA, Ph.D., Chair

KENNETH M. PORTIER, Ph.D., Session Chair

JOHN R. BUCHER, Ph.D., DABT

JANICE E. CHAMBERS, Ph.D., DABT, ATS

GERALD A. LeBLANC, Ph.D.

DANIEL SCHLENK, Ph.D.

FQPA SCIENCE REVIEW BOARD MEMBERS PRESENT:

SUSAN F. AKANA, Ph.D.

RICHARD H. COUPE, Ph.D.

KENNETH BARRY DELCLOS, Ph.D.

PENELOPE A. FENNER-CRISP, Ph.D., DABT

ROBERT J. GILLIOM, Ph.D.

RICHARD GREENWOOD, Ph.D.

WILLIAM L. HAYTON, Ph.D.

STEVEN D. HOLLADY, Ph.D.

TERESA H. HORTON, Ph.D.

KANNAN KRISHNAN, Ph.D.

HERBERT K.H. LEE, Ph.D.

KEVIN T. O'BYRNE, Ph.D.

NU-MAY RUBY REED, Ph.D., DABT

JEAN F.L. REGAL, Ph.D.

DANIEL J. SELVAGE, Ph.D.

CARMEN J. WILLIAMS, M.D., Ph.D.

LINDA J. YOUNG, Ph.D.

ALSO PRESENT:

JOSEPH E. BAILEY, Designated Federal Official

C-O-N-T-E-N-T-S

P-R-O-C-E-E-D-I-N-G-S

8:34 a.m.

MR. BAILEY: Okay, this is the last day of the FIFRA Scientific Advisory Panel meeting on atrazine reevaluation. I am Joe Bailey, serving as Designated Federal Official. Dr. Portier is the chair and I will turn the mike to him.

SESSION CHAIR PORTIER: Good morning. Thank all of you for sticking with us into this fourth day of discussion. Hopefully not a full day of discussion, but it remains to be seen.

Just to give the panel and the audience a little kind of idea of what our plans are this morning, I am going to revisit Question 1.9, just to get any concluding remarks and kind of make sure that at least the panel is clear that all of our ideas have been captured.

We will then go to some presentations from Syngenta and EPA on clarifying comments related to the hydrology and the simulations that they did. The panel members assigned to those

questions have been working very hard the last three days to really understand what both Syngenta and EPA has done so that they can make sure that their remarks are to the point.

And while it has been nice for them to have these discussions on the side, those discussions need to come back into the full room so they are captured into the public record. That is the gist of a public meeting.

So we will have those presentations hopefully no more than a half hour with questions and then we will begin with the four hydrology questions. And at that point, hopefully we will be able to close the public meeting and begin to write our report.

So with that, I am going to, I don't know where we are on the slides, but we are going to go back to Question 1.9 which if you remember, these are the questions that relate to the risk assessment and the PK and primarily deal with issues of frequency and duration of water monitoring as it relates to toxicological

concerns. And I have asked Dr. Reed and Dr. Krishnan, I warned them that I was going to call on them first to start the discussion and then we will open it back up to the panel.

Primarily, we don't really want to rehash everything we talked about yesterday but we want to capture any additional thoughts you may have had over the evening. It was an intensive two hours yesterday afternoon and I had to go back and think about a lot of it myself.

So with that, Dr. Reed?

DR. REED: Yes, the Chair gave me the warning but it was five minutes ago.

(Laughter.)

SESSION CHAIR PORTIER: You don't get much more for tornadoes, either.

DR. REED: Yes, for 1.9, the first part of it, we talked about let me see. I was given 30 seconds.

We talked about focusing on downstream events to define the endpoints for risk assessment. So these endpoints have to do with

something that is adverse in terms of function or anything that you would define.

And we talked a little bit about within the mode of action that we have been talking about in the last three days but also any possible other mode of actions. I have asked if anyone who has a sense of sensitivity difference between animals and humans of the particular endpoint that you are interested in having used risk assessment.

But most of the comments are really just to go back and look at the entire database and benchmark those analysis to form endpoints to compare the sensitivity of these endpoints and to include all the steps in the key events and mindful about the acute and short and long-terms duration of exposure and mindful about the toxicity of the metabolites and so some form of toxicity equivalence factor could be applied to it to address all the speciation of the chemicals.

And our sort of comments or inclusion

for this issue would be also regarding the lack of data from all the previous discussions.

And Dr. Horton had graciously offered to come up with maybe more detail. When we talk about key events and I felt yesterday we were a little bit ambiguous about what are the key events, how many, four, eight, or additional mode of action. And so Dr. Horton graciously offered to come up with a new house and maybe more elaborate.

So this will be where we want to hear from her. Dr. Horton.

DR. HORTON: Now I know why I didn't go into architecture.

Okay, I am still working on this new figure but I can give you an idea of what it will look like for the minutes. And it will be a multi-part figure reflecting the charge to the panel to evaluate the MOA and aid the agency in its preparation for the September 2010 meeting.

And the goal of agency by 2010 is to develop a draft weight of evidence document that

includes points of departures for evaluating risk in infants and children, with the goal of determining the extent to which the current data indicates a need for the Agency to develop a new human health assessment for atrazine and to reconsider, as appropriate, the frequency of drinking water sampling.

So to that end, the figure will be a multi-part figure encapsulating the mode of action or the modified mode of action as discussed here, which will be a newly conceptualized mode of action, taking into consideration the data discussed in the last few days. The weight of evidence supporting the various level data that has been discussed using arrows of different weights to indicate the confidence in the various pieces of evidence that have been discussed.

And the diagram will recognize that organisms, especially humans, are complex systems and that the systems we have discussing often interact and function as feedback systems with

some degree with hierarchical organization and that within that context, atrazine may act at several different levels of the hierarchy. And the key events may occur at any point within that hierarchy. And because these are feedback systems, they may ramify at different levels.

The diagram will also attempt to identify where new areas of research may be helpful or where areas of research need to be strengthened. Thank you.

DR. MENDEZ: Could I just one second please?

SESSION CHAIR PORTIER: Yes, sure.

DR. MENDEZ: I just want to clarify for the panel that our intent for September, it is to reevaluate the toxicity not just for infants and children but for the entire population.

SESSION CHAIR PORTIER: Sorry, this is Ken Portier. Is this on? Yes, I guess so.

In talking with Dr. Horton about the diagram, the one thing that caught me is she took

figure three and she said but the way I see it is this way and with CNS at the top and functional at the bottom. And I think that is one of the major changes.

It sounds simple but when she talked about top down or something at the top, you know, I was looking at it as atrazine from this chart and she was looking at it as CNS, looking at it in the transposed way. And to me, that clarified a lot of the conversations she had been making all along because she sees it in this rotated version with the central nervous system at the top and functional capabilities at the bottom and atrazine comes in from the side.

Dr. Horton and then Dr. Crisp.

DR. HORTON: You have to understand that within the world of reproductive endocrinologists and neuroendocrinologists, we divide the world into those of us who work above the belt and below the belt.

(Laughter.)

SESSION CHAIR PORTIER: Dr. Fenner-

Crisp?

DR. FENNER-CRISP: I don't know how to top that one.

I might suggest that you go back and look in the draft document with respect to the mammary tumor MOA and the figure in there does in fact start with the brain and trickle down.

SESSION CHAIR PORTIER: Dr. Cooper?

DR. COOPER: I would like to, if I could, also make a comment. I think that figure that you saw there was before it got to Washington.

But I am really pleased to hear you make the suggestions that you made because it is an extraordinarily complex set of issues that we are dealing with and the key point, you have captured it, especially when you think of toxicity pathways that was touched on yesterday, they are not linear. That they are probably masses with multiple, perhaps multiple target sites. And to capture that is an extraordinarily difficult thing, especially in a mixed audience

like this.

The other thing is that the arrows are key to this also. And there is a considerable amount of difference in the weight of evidence. But the only thing is, is that we are preparing, you are going to make suggestions about the way the arrows, the thickness of them is that in this case, I think there you can bring in some of the in vitro data that is existing.

SESSION CHAIR PORTIER: Dr. Bucher?

DR. BUCHER: Yes, I just wanted to bring up the fact I have been involved in a number of discussions over the years on MOA and the development of frameworks for the utilization of MOA with respect to evaluating human cancer information. And one of the things that I think is most important is that the MOAs have to be pretty well developed on, I think the words we have used are compelling, scientifically compelling.

And what I am a little worried about here is that we have a very, very complicated

potential figure for an MOA that is being introduced at the very end of the meeting and I wouldn't want it to come out of this meeting as something to the world that has been fully evaluated and agreed upon by this group.

So I think that whatever you put together should be cast in the right terms that it is a very, you know, it is a suggestion of a way that the agency might want to look at organizing the information rather than being cast as an MOA.

SESSION CHAIR PORTIER: Yes, we have talked about that as well. Dr. Krishnan, do you want to talk about PK issues?

DR. KRISHNAN: The second part of this question related to the temporal consideration based on toxicity versus monitoring frequency. As much as I would like to talk about PK, I am going to have to be a little bit more implicit of both PK and PD. That is what I am going to try to do and invite Dr. O'Bryne to complete or correct me as I make my comments.

In thinking about this, you know, frequency of monitoring versus the temporal profile or temporal considerations of toxicity, we do it in terms of the time course of the atrazine and metabolite in the body, which is essentially the PK consideration, and then the time course or the precursor, the key events or the precursors measured, which could be altered LH cortical levels.

First, in terms of the consideration of the time course of atrazine and metabolites or the internal dose. As we talked about yesterday, the uptake or consumption pattern associated with the drinking water combined with the pharmacokinetic considerations, my colleagues mentioned some of the key aspects, particularly the slow absorption rate for example from the drinking water or from oral administrations, the relative short half-life of the parent chemical combined with an extended half-life of some of the metabolites, these things suggest together that in the integrated measure, the internal

exposure would be relevant.

For example, the chloroforms, the chlorinated form, atrazine plus metabolites that still have the Cl in them. Or there could be an average calculated based on the AUCs over a particular period.

So the area under the curves would integrate both the dose and the temporal considerations together, essentially. And I have not seen anything that suggests that the particular Cmax of a parent chemical or metabolite during a specific period is somehow clearly associated with the profile of the precursors measured or the outcome in tox studies.

So given those observations, I still tend to think that the overall profile or consideration of pharmacokinetics does not call for very narrow time-based analysis of monitoring. So that is one part of it.

The second part is that then you think about the time course of the key events are

essentially the precursor or the measurements that are made between the internal dose and the toxicological outcome. In this case, the time course of the LH or the cortical levels.

The current complication, in my mind at least, arises from the consideration of the more recent data on the short-term changes in the rats summarized in Table 3 on the HPA/HPG axis, essentially in the four-day experiment. I think that is where some of the discussions yesterday focused on.

It is kind of unclear to me if the four-day exposures result in an impact on the precursors, for example, LH or cortical levels. And since they seem to bounce back within the next few days that follow, these four-day experiments which lead to all the discussions of a four-day monitoring frequency, for example, is kind of related to an adapting response, rather than an adverse response. That is a question I ask myself. You know, if it is a four-day dosing that causes an impact on LH and then in four days

it bounces back, then the cycle, I guess, it gets back afterwards and so forth, it appears to me more of an adapted response, rather than adverse reproductive outcome per se.

Further, depending on the critical effect, I mean, that is one of the reasons why we revisited the MOA figure. Depending upon the critical effect that is going to drive the acute versus chronic assessment, I hesitate to see this single MOA funnel in accommodating all of these outcomes. Specifically, I don't think one MOA will fit all effects acute and chronic. That is where some of my concerns come.

So the precursor effects appear to make at least convincing sense to me based on chronic exposures and sustained effects on the LH associated with the reproductive outcome and so on which are not questioned at all.

But in terms of the acute exposure and I am still a bit reluctant to suggest anything other than an integrated or an average one, rather than focusing on a four-day period based

on these points.

Additionally, as I was thinking about this four-day discussion yesterday and some of it came through during the supper time, is that you know, once you take a benchmark, some level of a NOAEL in a rat study, based on the effects on the precursors and then use that as a basis for deriving oral concentrations appropriate for humans, what is the point of relating then back to the four-day frequency there or the duration of the effect in the rat unless we want to protect the rat? Because that is where some of our confusions came from.

The focus should be on the dose because we took the dose based on the rat studies and we said well we will use that as a basis to protect the humans. And then you do the calculations and then I don't see the obvious connection of going back with these numbers, back to the four-day frequency. That obvious link, I mean, that is not an obvious link to me, unless you want to calculate the physiological equal and

times and so forth. Even there I am not sure you would say it is a worthy effort.

As Bob Dedrick, I think, in the '70s derived some of these physiological time equivalents and so forth, those sophisticated calculations could be done but I don't see that being a productive route.

So, I don't know if I confused or clarified more but this is what is in my mind, based on discussions we have had.

SESSION CHAIR PORTIER: Dr. Greenwood, you wanted to add something to this?

DR. GREENWOOD: Yes. I think we were trying to get our heads around this at the end of the session yesterday afternoon. And I think what we saw was that it makes a lot of sense, because you have got to base your protective levels on the hazard. And we saw that there was a lot of sense in basing it on the rodent assay, which is clear, with a clear endpoint and so on, and well supported with evidence.

But we couldn't see then why the time

should be based on the rodent at all. I mean, this is what Dr. Krishnan just said. We just could not see why you would use a time based for a rat effect. Okay, the dose is important. That is what you use it for, to get the hazard. But I don't think it was ever intended to try and set an exposure time. As Dr. Krishnan said, you are not trying to, a lot of people spend money trying to kill rats and not trying to protect them.

So I think we felt that if there were to be an exposure time figured in to the water monitoring, it has really got to be something more human-based. But what that would be is difficult because, as Dr. Horton said, it could be just an hour, the critical period for some endpoints and you are never going to capture that.

So we felt that if you go for the most vulnerable stage, maybe during development, it could just be an hour but I would prefer to let Dr. Horton speak to that, I think.

SESSION CHAIR PORTIER: Dr. Horton?

DR. HORTON: Yes, this is one of the things we will try to encapsulate in the new diagram is the difference between developmental programming effects, which may have long-term outcomes which may not, while exposure may occur during fetal development and may result from short duration exposure but the actual effects may not appear until later in life versus what Dr. Krishnan referred to as an adaptive response occurring in response to an acute exposure. So we will try to capture those two different things because in a physiological sense, the mechanisms of action may be very different.

SESSION CHAIR PORTIER: Dr. Bucher?

DR. BUCHER: So I understand and I agree with the comments you made about the time issues. But I think that you said something about the fact that if there was a four-day exposure and there was an effect but it was reversible, it wasn't considered an adverse effect. And I would absolutely disagree with that.

And if you are talking about an adaptive behavior as simply something that can be reversible, I would disagree with that as well. I think that is not the correct interpretation of toxicology.

SESSION CHAIR PORTIER: Dr. Cooper?

DR. COOPER: This is not an atrazine paper but in 1993 we published a paper with the acaricide chlordimeform where we showed that a single dose administered on the critical period on the afternoon of proestrus and the clearance of this stuff is pretty quick, blocked the LH surge and delayed ovulation. And everyone said well that is a reversible effect and if you block ovulation, they ovulate the next day, especially a rat. And rats are rats and humans are humans.

Well, it turns out that we went on to examine that further. And when you delay ovulation, the ova ages. And as the ova ages, it becomes, I don't know what the right word is, the viability is such that you end up with things like polyspermia and you end up with polyploidy.

I think that was Dr. Stoker's master's thesis.

And so when you went on and looked at the mothers of those one-time dosed animals which was reversible, if they became pregnant, then that effect was not reversible. There was delayed development, embryo anomalies, reduced litter size. So sometimes a single dose, even though it may apparently be reversible, may not necessarily be that.

SESSION CHAIR PORTIER: Dr. Krishnan?

DR. KRISHNAN: I think I was cautious in the way I phrased it, but what I had raised as a question or concern was that those effects during the four days on the precursors if you will, the cortical and the LH, if they are clearly associated with the reproductive outcome or the delayed maturation on some of the other whole animal affects, clearly, then I don't ask question aloud. So it is obvious.

My understanding or at least the reason why I raised the question was if it was affecting a cycle on then which bounces back and

then captures, there is no demonstration of reproductive outcome or other whole animal effects in these studies, which doesn't seem to be the case. If not, please correct me. Then I will take back and I say well three or four-day or even less or more frequent monitoring would be required.

So I just want to be clear that I have not misread or misinterpreted.

SESSION CHAIR PORTIER: Dr. Reed?

DR. REED: Just so that you are having fun here, when I was beginning to do risk assessment and our kids were very young and they asked me what do you do and so I said we just do experiments on rats and mice and bunnies and then we bring that down by hundred-fold. And they got the idea that all the field mice and field rats and bunnies are all very safe.

And about the reversibility, I totally agree with Dr. Bucher that reversible effects are still adverse if it is adverse to begin with.

Also, a lot of times maternal effects

may be reversible. Alkaloids is an example that the fetus is not, when it gets to the fetus effect, fetal effects.

But I think, my understanding of this issue about reversibility is that you have a cascade of events and going through the networks or the pathway of the mode of action, there may be a step that you would have effect that there is some sort of threshold so that it will not trigger the next step, then that might be something to consider as "reversible" in that context.

But my main comments is that we talk a little bit about area under the curve and blood concentration and so forth. But I think it is important, at least when I am doing PAPK model, the hardest thing is to decide on the dose metrics for rats to human equivalents. And I think that when we get to PAPK or use PAPK as a tool, that we should clear about what is the most appropriate target sites. And I would not preclude the possibility of using peak

concentration and just across the board to say area under the curve is the most appropriate. In terms of target site, it might not be serum but could be more pertinent. Could be brain.

And so these are the kind of things that I think we need to consider. I don't think it is necessary for us to set in stone to say area under the curve or what target site, or just serum concentration.

SESSION CHAIR PORTIER: Dr. O'Byrne?

DR. O'BYRNE: I am just listening to these conversations, I am trying to remind myself why I am here. And it was my thought that we had this new data of this 15-minute activation of the HP axis, which was making us reconsider the sampling frequency.

Now, my considered opinion is that this 15-minute HP activation is a red herring and it reminds me of Tony Blair's 45-minute weapons of mass destruction argument. And I think we have got to be very careful here.

The data, I mean, we have heard these

arguments about development and how sensitive they may be but I haven't actually seen any data. We are dealing here with a surge as this sort of benchmark which cannot be separated from cycles. If you don't have surges, you don't have cycles. And the rat has got a four-day cycle and humans have got a 28-day cycle. And the data that we have got in front of us is that the LOEL for atrazine is 3.6 milligrams over 28 weeks or 6.2 over four days. And that is what we have to work with. So, I think these are the critical factors.

And the discussions about how to extrapolate from the rat model to humans is also quite important, I think, to appreciate. And to fit the physiology into the toxicology is hugely difficult and we struggle with that. And I don't know whether you can extrapolate from some analogies and experimental data.

There is one that comes to my mind. If you take a rat and you give it a large enough dose of estrogen, then you will have an LH surge

every day at 5:00 or whatever until the thing becomes exhausted.

If you give a large dose of estrogen to a woman at any stage during the follicular phase, you stop that cycle dead in its tracks and you will wait 14 days for the next spontaneous LH surge. So I don't know whether you can use that temporal concordance or not but it is hellishly difficult.

I don't know if that is of any value at all.

SESSION CHAIR PORTIER: I am sure the Agency is going to have to deal with that as they kind of move forward and put this all together.

I am going to kind of end it at this point. I just wanted to make a last comment that as I was thinking through this and trying to make the transition to the monitoring and sampling, while it is nice to understand the mechanism of action and understand this stuff, when you really think from a population public health level, you know, you start thinking about a hundred million

women, at any one point, any one day, any one point in time, some of them are at the critical level. So the statistic you are probably looking for for monitoring is going to be some kind of daily average that we are not going to exceed because there is somebody at risk every day.

So it is nice to know the duration of the impact of a concentration on an individual but when you integrate it over the population, it is going to be an average value that you are looking at. The question is translating that average to a physiologically critical dose that is safe. Right? And that is the concern.

So I hope, Dr. Krishnan and Dr. Reed are going to be able to capture this. The reason for doing this is I know this is a critical question and I really want to make sure as we write this up we kind of capture all of this.

Thank you. I think, Dr. Cooper, you are done. We are going to bring up the hydrology point.

At this point, I think Syngenta is

first up. Dr. Hendley, I think has the microphone, followed by Dr. Sielken, followed by Mr. Thurman or Mary Frankenberry. And they promised me to kind of collectively keep this within 30 or 40 minutes. And we will probably, at the end of this presentation, we will probably take our morning break, get our coffee and then go into these questions.

A number of pages have been handed out to the panel and they are already being processed to put on the docket, some of which are going to be presented here and some of which have not. A lot of it is clarifying material.

We went through the presentation on Monday afternoon kind of quick and that was, I think, the one reason we really want to revisit and make sure we understand what was presented. I know it is early in the morning but I am sure the toxicology physiologists over here are going to kind of sit back and relax.

Dr. Hendley.

DR. HENDLEY: Okay, this is Paul

Hendley from Syngenta again. And we would like to thank you, Mr. Chairman and the panel for giving us an opportunity to clarify some of the handouts that were given.

And I think in fairness to the hydrologists, for the rest of you around the panel, you need to realize that between ourselves and EPA, we have talked about five approaches, each of which have examined data in subtly different ways than the hydrologists have properly asked us to clarify exactly how the data were processed so they can better understand the interpretations that have been made.

So what I am going to be doing is referring or what we are going to be doing, and incidentally, I have Dr. Sielken on my left and Dr. Chen, the Dr. Chen from Chen et al., which you have seen a number of times, on my right.

What we are going to do is talk about the slide number from my presentation and then, if I may, Mr. Chairman, I suggest after these are largely one pages, if we briefly go through the

one-pager and then ask to make sure we have addressed the questions, it is more efficient than trying to remember what we said. Thank you very much, indeed, then.

Okay, the first one refers to --

SESSION CHAIR PORTIER: Can you make that full screen?

DR. HENDLEY: I can. I can make it bigger but I am afraid it is in Word. Where is the zoom on this version of Word? It is not the one where -- zoom. Thank you. Is that any better? No, not really. Thank you.

It is amazing how Windows changes. You totally forget the version as you move on.

Okay, the first one refers to slide ten in the Syngenta presentation, which was about high centiles. There is a very detailed report on the docket, which is Whitmore and Mosquin.

The points that I think I would like to clarify that I understand were questions in the minds of the hydrologists were this is finished water. And what we are talking about

here is an analysis that follows almost exactly the process in the main report, which was done on the full data set of both SDWA 50,000 data points and VMP/AMP, the 48,000 data points. But this time, they looked at the subsets of data based on the water sources, static, flowing, and mixed.

And in addition, they looked at the subset of data based across all community water systems but recorded samples between April 1 and July 31.

And so briefly, in that process, each subset of the VMP/AMP data set was taken. The data points were weighted, as described in detail in the report. And we probably don't need to go into that today.

The key question I think that was in the mind of the panel was were these data points interpolated. And the answer is they were not interpolated. The centile estimates that were given on slide ten were computed directly from the population of measurements. And then the confidence intervals were obtained as described

in the full report and that led to the table in the handout that accompanied the slides.

DR. YOUNG: Linda Young. Okay, Paul, I can't figure out, are you talking about this handout, and this slide ten? I can't find slide ten.

DR. HENDLEY: Okay.

DR. YOUNG: So, I am quite lost.

DR. HENDLEY: Okay. I do not have the slide set up here. It is the slide set that I presented. I think it is near the back of that slide, that package of slides.

DR. YOUNG: Okay.

DR. HENDLEY: There you go.

DR. YOUNG: All right.

DR. HENDLEY: No, the slide set that starts with potential atrazine exposure in drinking water. You got it. That is slide ten.

So that was on the high centiles and it had the data on the full data set, as well as the subsets by source type, static water, the flowing water. And so that was done with no

interpolation from the entire set of data points. Any further questions on that one? Okay.

The second one refers to slide 12 in that presentation, which you should find says sampling frequency and then 90-day exposure. Okay? Right.

So, this is a handout that is describing in more detail the process of developing the data. And the key points here, it was raw water and there were 441 community water system years from -- and now we can read it -- from 137 community water systems.

This was from the atrazine monitoring program. So they were seven-day intervals during the growing season and every two weeks during the rest of the season. And for each one of those 441 community water system years, linear interpolation was used in this case to create a 365 daily concentration profile. From that interpolated profile, 90-day rolling averages were computed and the maximum one was selected for each community water system year. That was

the true value against what we are going to compare our samples.

There were a number of sampling designs tested and just to be clear, they are as shown on the table on the other side. Overleaf, as the Brits would say. This is the atrazine monitoring protocol, right to extreme right of that table. And in January, there was one every two weeks, which works out to approximately three sets of samples in January, two in February, two in March, and this just happens to be because of calendar weeks and then you go into the weekly period.

SESSION CHAIR PORTIER: Time is a-wasting here.

DR. HENDLEY: Yes. Okay. And so we had weekly sampling during April to July and then again went back to the two-weekly schedule for the rest of the year and there were also sampling regimes with 17 samples, nine and six during the year.

The sampling design that was used in

this piece of work by Research Triangle International was a series of windows were created and the approach was to take one sample in each simulation, one sample point within the window with replacement. Each day in the window had an equal probability of selection for all 1,000 simulations.

And so you took a random day within the week. You moved to the next selection week, took another random day. And the reason for that is because that is actually the instruction that was given to the plan operators, not that they had to go every Tuesday, because that was an unfair burden but they were to take one during the calendar weeks, following that schedule in the table.

So for each simulation there was a linear interpolation between the sampled points to come up with a 365-day profile and from that, a 90-day rolling average was created for the days and the maximum simulation 90-day rolling average was obtained.

And the difference between the simulation 90-day maximum and the true maximum for each of those community water system years was recorded and then you had a set of maximum simulated to maximum and that was worked on as a full data set and as data subsets based on the source-type. So that was the background to how we processed the data for that piece of work. And I realize you need many hands to hold all the pieces of paper.

DR. YOUNG: So the results are in this set of slides?

DR. HENDLEY: The results are summarized on slide 12 of the presentation. The 90-day exposure slide, which is entitled "Sampling Frequency and Confidence; 90-day Exposure."

DR. SIELKEN: This is Dr. Sielken. There is an analogous set of answers in the other handout that you had that I provided that was done using the procedure that I am going to describe later. But there are also additional

numerical results for that type of analysis.

DR. HENDLEY: And of course, the full background to this, Dr. Young, is in the Chen, et al. 2009, the evaluation of sampling frequency alternatives, which is in the docket.

SESSION CHAIR PORTIER: And I think the key conclusion here is that when you use that methodology to look at reduced sampling schemes, so going from roughly 25, 26 samples in a year down to six or nine samples and showed that by that same methodology you got the same distributions with fewer samples.

DR. HENDLEY: Precisely. Precisely. Thank you. Any further questions on that one? Okay.

Okay, the last one of my presentation before Dr. Sielken refers to slide number 14 in the presentation that I made and it is entitled "Sample Frequency, Shorter Exposures." This one will be quicker, Mr. Chairman.

The key points of it, unlike the last one where we started off with atrazine monitoring

program frequency, which was basically seven days apart, this is using daily data sets. And I should say, there is one other data set we haven't been specific about and Dr. Gilliom can tell you more about it, and that is a set from 2001 for Lake Perry, which was done by Blomquist, et al., it is submitted, it is in the EPA White Paper and that had daily data for a large static body. And I realize that we never actually mentioned that here.

So as you remember, we had finished data from a community water system in Missouri that was daily or near daily. And we had daily or near daily data from the Eco program and the Heidelberg programs.

So for each source of data for each year, we took the measured daily or near daily data. We filled in any gaps with linear interpolation to create a 365-day time series and we identified peak and we computed point estimates of the 95th and 99th centile daily value. And those were the true ones against

which we compared.

DR. YOUNG: So how were those percentiles estimated?

DR. HENDLEY: Dr. Chen?

DR. CHEN: This is Wenlin Chen, Syngenta.

The percentile actually are calculated based on the daily chemograph generated. We looked at the maximum each year at each site. You have a sequence of events and you keep the maximum as a peak concentration for that year for that site and then calculate those so the 90th centile and the 99th centile.

DR. YOUNG: So you used, basically the rate data and took the percentiles of the data?

DR. CHEN: Right, yes.

DR. YOUNG: Okay.

DR. HENDLEY: So again, using that 365-day time series for that source of data for that year, we used the same simulation approach using the window approach that we have just discussed. And so for each resulting simulation

of those one thousand simulations, a linear interpolation between the sample points generated a 365-day profile as before and the peak and 95th and 99th centile of daily values were obtained from that.

And then from each distribution of a thousand runs, the example in the handout that was given, this was from Dr. Mosquin, shows the mean, the peak, and the 95th centile daily values and their percentage deviations.

So has that clarified the data processing for that?

DR. SIELKEN: This is Dr. Sielken. Dr. Mosquin's handout that he is referring to are the two separate pages that were put in the docket from Dr. Mosquin?

DR. HENDLEY: Yes, the last two in there.

SESSION CHAIR PORTIER: And the conclusion from this, again getting back to the conclusions of this exercise, is that when you compared a once-a-week sampling to your overall,

you tend to underestimate the maximum by 20 percent, 20 percent to 80 percent for Missouri, and 1.1 to 2.6 multipliers for the others. So as we would expect, you are underestimating the maximum. What you were doing here was estimating the multiplier between the maximum between the sample distribution to the real. Being able to say, if I had sampled once a week during the season and used that to create my profile, I would underestimate the maximum by 20 to 80 percent. Right? I mean, is that kind of the conclusion?

DR. CHEN: These systems are not -- sorry. This is Wenlin Chen. You mentioned the three systems. One is St. Louis which is a true drinking system, the other two are at the Ohio River and the eco that is not drinking water. That is just ecological water system.

SESSION CHAIR PORTIER: And you get the performance with the drinking water profile and the worst performance with the well water profiles, as we would expect.

Dr. Heeringa?

CHAIR HEERINGA: Steve Heeringa. Paul, there is a chemograph that is illustrated on this handout sheet. Is that typical of one of these water systems? For example, you are using site-specific, year-specific data. So you have seven years for this Missouri finished water system. What do those chemographs look like?

DR. HENDLEY: Paul Hendley. The Missouri chemographs have a low maximum value but they are actually quite spiky. In a way, quite surprisingly spiky. So some of the peaks appear to be of the order of maybe three or four days.

DR. CHEN: Wenlin Chen. I just wanted to add to that is that we are looking at the ratio, not the absolute of the peak. So when you look at the ratio, it should give you, it cancels out where it is really spiky or not spiky.

CHAIR HEERINGA: Steve Heeringa. My concern is that the spiky problem is the tough one. And I assume the Heidelberg data, which are Honey Creek and Rock Creek, that those are spiky

data as well because those are agricultural drainage, mostly.

DR. HENDLEY: Especially for the two that you have selected, which were Honey and Rock, which were about 35 and I think 130 square miles.

CHAIR HEERINGA: Thank you.

SESSION CHAIR PORTIER: Paul, scroll down so people can see the image Dr. Heeringa was talking about. I think it is further down on your slide. Isn't it?

DR. HENDLEY: I think it is on one of the others, yes.

SESSION CHAIR PORTIER: There it is.

DR. HENDLEY: There you go.

SESSION CHAIR PORTIER: And the red lines identify the sampling bins, the periods for the windows.

DR. HENDLEY: Absolutely correctly. So what you are doing is you are getting the computer to pull a sample from each of these windows.

So with that, if there is no further questions from the hydrology group, we would like to turn to the discussion for underlying -- Where is your presentation? And this underlies slide 13

DR. SIELKEN: This is Dr. Sielken. Thank you, Mr. Chairman and the rest of the panel.

This is some supplemental explanation to go with the short handout that was part of the docket yesterday. There seemed to be a missing link in the middle of it. And since it was no simulation and no trickery or anything like that, I wanted to make it clear the step-by-step procedure that I was going through and also differentiated a little bit from some of the -- slightly different than the Mosquin procedure, although our procedure and our results ended up being within ten percent of each other. So they were very close, even though they were slightly different.

The procedure that I applied both to

all 202 CWS's as well and that was on a multi-year profile, I also applied to the St. Louis -- I'm sorry. I wasn't supposed to say that. The Missouri community water supply system, which was close to daily data. I did that both yearly in groups and multi-year. And I also applied that to, the same system, to the almost daily data in the Heidelberg data sets, the Honey, the Rock, Sandusky, and Maumee.

In all of those cases, the procedure was as follows. And it is almost identical whether you use linear interpolation or step-wise. So that is not really a big deal.

You start out, at least for data sets like this where you followed this nearly weekly or nearly daily sampling. It is a fairly dense data set, as shown on the top of this slide. The slide underneath it is the linear interpolation and that is where we started. And the only modeling that was done was just that linear interpolation. There was nothing else that was simulated.

SESSION CHAIR PORTIER: And for those who can't read it, that is about 13 years' worth of time series data for one site that you are showing there.

DR. SIELKEN: Yes, that is correct. We had a variable number of years in CWSes. Most of them had around ten years. Well, there were 351 profiles that were 13 years long; 351 CWS years came from 13-year profiles. So that was 13-year profiles and seven-year profiles were the most common length of profiles.

From this linear interpolated profile for multi-years capturing your variability, I went through and overlaid a grid of seven days wide. A window, a seven-day wide window. And then picture of course that seven-day window is not to scale. But you know, put a grid over it of seven days and then started at the beginning of the profile on Monday, went to the next Monday, took another -- and this is in the seven-day sampling.

We actually did the testing because

you were interested in perhaps changing the frequency of sampling. We also looked at two-day, three-day, four-day, five-day, six-day, seven-day, weekly, bi-weekly, tri-weekly, and monthly. So we looked at all of those profiles in exactly the same way that I am going to show here and I will use as an example the seven-day spacing. So analogous to what is going on now.

Given the linear profile, we overlay a grid of the sample spacing here seven days apart. Start systematically with Mondays and then go to the next Monday and so forth and just march through that entire profile every seven days, starting with a Monday, record those values, take a linear profile between those seven day values -- no. Sorry.

Take those seven day values, find the max, and then compare that max to the original profile max, the data profile max. So when we are looking at acute for one day, we just take those seven-day spaced apart samples, compare the max in that set of samples to the max in the

original profile. So there is no simulation in there.

I did that starting on Monday, starting on Tuesday, Wednesday, Thursday, you know, all the seven possible starting points. So there were seven sample maxes to compare to the overall data profile max. And you can see the variability in those numbers on this profile. You know, 0.97 for Mondays, 0.96 for Tuesdays. Obviously, one of those was going to hit it and it was Saturday in this example, sort of.

I took the average of those seven numbers and that is what I recorded as the single number for that CWS is a ratio of the sample max to the data profile max.

And now this next slide is the new one. And that was to say what did I do with those 202 numbers? Well, first of all, that is what I worked with was those 202 numbers; one for each CWS, just a comparison of the sample max to the profile max. I took those 202 numbers and found the percentiles of those 202 numbers. So

there was no simulation of anything in there.

And the little histogram here in the center is just to hopefully eliminate confusion. If there was a grid, and all of these ratios for seven day spacing and trying to target an acute daily value or a daily value, the ratios between the sample max and the true max ranged from 0.75 at the low end up to one. There was 18 from 0.75 to 0.80; 31 in the next bin, and so forth.

So they were not very widely spread. The worst case was around 0.75. So you were only off by a quarter, 25 percent, using seven-day sampling to estimate the ratio between a max one day and a profile one day. So for a one-day target, you can use a seven-day sampling and be within 25 percent.

Dr. Lee?

DR. LEE: For the ones where it is you are sampling say every seven days, did you average across the possible seven starting days to get the figures in the table?

DR. SIELKEN: Yes, I did.

DR. LEE: Okay, thank you.

DR. SIELKEN: So, it was the expected performance of the sampling plan.

Okay, that was probably the slide that was most missing from your explanation before.

That set of percentiles computed from the 202 values was tabulated in your handout for both one-day sampling, of course that would have been perfect; two-day sampling; three-day spacing; all the way up to 28-day spacing. So this is just a table repeating the analysis for each sample spacing and then just recording the percentile.

SESSION CHAIR PORTIER: If I can interpret this and then if you look at the last column, 28 days, it says if instead of sampling every day you sample once a month, kind of on average you are going to whatever it is, 22 percent below the kind of in the long run, you are going to be 22 percent below the true daily maximum. Right?

DR. SIELKEN: I'm not exactly sure

where you are seeing 22 --

SESSION CHAIR PORTIER: I am looking at the 50th percentile. I'm sorry. The median.

DR. SIELKEN: Oh, at the 50th percentile. Okay. Going right across here, 0.7548, you are about 25 percent below.

SESSION CHAIR PORTIER: It is 0.79. So it is like 21 percent.

DR. SIELKEN: Well okay, yes.

SESSION CHAIR PORTIER: You can't read it from there.

DR. SIELKEN: Knowing that tables are hard to read, regardless of the magnitude, I put a little picture in and you can't read that either but it gave me the idea that as you varied the sample spacing between one day, which gave you perfect results, to 28 days, how did your ratio drop off at the 95th? Well, the orange, if you can see orange is in the middle, and that is the 95th, there is the 97.5 and 90 below it. So those three profiles of interest, how they change with sample spacing.

And notice that if you don't like pictures, I put it in numbers again. And they are saying that if you take a seven-day sample spacing, keep your current spacing, you are within a quarter for single-day target within 20 percent for a three-day target, and so forth. Notice that for a 90-day rolling target, you are almost there all the time.

So this is showing that regardless of your target, really, you are doing quite well with the existing plan. Now this was to test the sampling performance.

What happened? Oh, well, I reached -- well how silly. I want to get back to the folders. Where is the folders? Yes, here we go.

That showed that you could stick with the current sampling plan and for finished water or even doing the same example with the Heidelberg Eco, getting approximately comparable results. And those were shown in the first handout that I gave, put on the docket a couple of days ago down here. I apologize here.

We did that same procedure as Dr. Hendley mentioned, we did that not only for the 202 CWS collectively, we also split them up by groups, flowing, mixed and static. We got pretty much the same results. You can see a comparison there between what we got for all the CWS together with what we got for the different partition. That is in the handout that you got on the first day from me, which is going to look like that.

And then because this was the 202 CWS, which is a really strong database for capturing year to year variability and the differences between CWS in Indiana, Illinois, Ohio, Texas, you know, covering that entire region and the different year effects.

And of course, if you look at a CWS, there is a lot of year to year variability. Then going from that set as a test case to another set that didn't hardly involve the linear interpolation because you had almost daily sampling, which is the CWS set in Missouri, which

is shown in this slide. Yes, well if I do that then I lose where I was.

You can notice these numbers down here. The lowest number in that table is, you know, for the hardest target, which is single day, those numbers are about a third. The lowest number there is 65 percent or something like that, 60 percent. That means you are off by 40 percent, not even a factor of two.

Okay, also you know, this was for almost daily sampling of finished water. This is the Heidelberg data set. Again, the two worst cases and again, this is not drinking water but an ecosystem, your lowest number is around 40 percent. So you are off by that much. So again, right around at worst a factor of two.

DR. HENDLEY: And just if I can, this is Paul Hendley again, that was why having worked from the strength of the temporal and spatial VMP/AMP database and from the strength of the daily database and coming together and finding the results were coherent was one of the points

we made earlier.

SESSION CHAIR PORTIER: So do we have any follow-up questions of Linda or Dr. Lee? Dr. Heeringa.

CHAIR HEERINGA: I just want confirmation that the two sets of analyses, the one done by RTI and then the one done by Sielken and Associates, Bob you just used systematic sampling throughout the year. The other just used the random draw within these time windows that have been set out under the plan.

DR. HENDLEY: That is correct.

CHAIR HEERINGA: That is the primary difference between your treatments of these same data?

DR. HENDLEY: Yes, that is primarily the difference. The only other difference is they looked at every year individually. And I did, too, but I did a multi-year profile.

There was one other comment I wanted to make because you asked it yourself, Dr. Heeringa, was what durations. In the first part

of the handout, we looked at, you know, gave the full numerical distributions of characteristics for one day, ten day, 30 day and 90 day rolling averages. We gave the concentrations, the distribution of concentrations, the distribution of rolling averages. We also addressed the question that you raised was what about how many days in a row did you exceed certain levels.

And there was, in that handout, I will just leave it here, in that handout, there are tables that do for the daily profiles the concentrations and then there is just this one page on page five in that handout where we did look at duration, number of days above specified values. And I only have the one day and the ten day in here. But I can tell you that at the 99th percentile, there was no CWS, none of the 202 CWS for finished water that had durations above twelve and a half for more than a day. And so there was no duration above twelve and a half.

Obviously at the extremes there was a little bit, I mean, there was more duration at

the max but there wasn't, you know, at the 99th percentile, there was none for the day, 10-day, 30-day, 90-day, there were no durations above that target value of like twelve and a half.

SESSION CHAIR PORTIER: Okay. Yes, Dr. Gilliom?

DR. GILLIOM: I just want to make a general point that we will probably come back to in discussing the charge questions but it is important in just interpreting all of this. There is a tremendous number of numbers. And you can generate tons of numbers from synthetic sampling experiments and everything.

I just want to make the general point that the importance of each sampling experiment to a specific objective questions depends entirely on how well the simulated truth represents actual truth.

So, if we get to a problem where we are looking at a very short-term occurrence like daily, we have to be sure the starting point is a confident estimate of the true daily

distribution. And there is a lot and I don't even intend to jump into every individual experiment that has been done by Syngenta or EPA but every one of them has a little different twist on how truth was defined and what time span it is relevant to.

So my general point is, there are so many possibilities here we are going to have to get that objective very clearly defined so that we can then evaluate which one is the right one to use.

DR. SIELKEN: I agree with Dr. Gilliom that establishing the truth is an important thing. I would point out that when we did the, we took the 202 CWS and took the linear profile as a starting point, my intention was not that that necessarily captured the max within the water that was actually there but if I took those values which are a representation of reality, took that as a reality, it may not have been quite as much of a reality as some people would have wanted if they were to sample more often.

But with that as the reality, given that reality, how well did you do with your monitoring program? So that was really the issue was the ability or the performance of the sampling. And that is also why we turned to the more daily profile values and tested it there.

Thank you.

SESSION CHAIR PORTIER: Paul, you want to wrap up? Final comment? No.

Dr. Akana had a question.

DR. AKANA: A small point that you can clarify for me. What I understand correctly though, this data treatment is equally valid on say the raw samples that start much higher as well as lower. For instance, the raw data here mostly the points are like three or under parts per billion but the treatment is equally valid if your dataset runs up to 30?

DR. SIELKEN: Yes. Yes, because we were looking at ratios between the sample max and the true max, that ratio would be invariant to whether we were going zero to three, zero to 30.

You are absolutely correct. It would be applicable.

DR. AKANA: But that does mean if your estimate is say 20 percent under-represented, for three, it is 20 percent and for 30 it is 20 percent.

DR. SIELKEN: Yes, that is correct.

DR. HENDLEY: So Mr. Chairman, I would like to wrap up. I will make one comment on that. Of course, that is a correct statement that as we pointed out before, if you are trying to understand the variability for drinking water values, you are best off using drinking water data where it is all possible. But the raw data was a great way of getting a handle on understanding variability.

However, we do appreciate the time you gave us to try and clarify some of these issues. So thank you very much. We appreciate it.

SESSION CHAIR PORTIER: Thank you. And now Nelson Thurman and Mary Frankenberry. I always get her last name wrong. I just know her

as Mary.

And also at the table is Don Brady, the vision man. And of course we are way beyond my half hour target. This is the last half hour before our discussion.

MS. FRANKENBERRY: Thank you. And again, I can go quickly I hope. Our slides I think are a lot simpler than Syngenta's. We did not get their handouts until one of them this morning and would like to get them from Joe Bailey, I think, certainly before the end of the day.

Hopefully with these, it appears that they have done at least in one or two of their exercises something very similar to what we did and I am hoping that these will be easy to understand.

Step one, we took a sample chemograph. That is what we get from the field, 30 to 35 samples. We augmented it linearly to 365 days to make a true or reference profile that we are now calling true.

What we got in step two we would like to consider reality. It could have been instead the Heidelberg dataset that would have started with, we wouldn't have had to do as many interpolations but this is what we got and considered as true.

From that we sampled this true profile or what we consider reality out in the field. Let's just say we will look at the example of sampling every four days. We took this -- this is what we do in the field. We may sample every seven days, actually in the AMP program. For this example, try four days.

What we get from number three then, is what we get our as our sample dataset from Syngenta or from wherever we receive a dataset. And that will have 90 some values, perhaps or 30 to 35 if it is from the AMP.

For the purpose of the exercises then we augmented this new sample chemograph up to 365 days. In some runs we ran step-wise, others linearly but what we presented in the paper were

linear interpolation.

For that new, that sample that was interpolated from step four, we calculated, as an example, three-day running averages, starting days one to three, two to four, all the way up to 363 to 365 days. We did rolling three-day averages there.

In step six, we took the maximum three-day running average from this re-sample or the first re-sample, if you will. We set it aside into the bootstrap pile and that is what we will call it without the acronym there.

(Laughter.)

MS. FRANKENBERRY: I actually made myself laugh last night and then forgot about the public record.

(Laughter.)

MS. FRANKENBERRY: In step six, we repeated this 4500 times with replacement until our bootstrap pile contained four to five thousand maximum three-day running averages. So what we have there is a distribution of maximum

three-day averages, three-day running averages that was derived from sampling every four days. That is our bootstrap sample of maximum three-day averages.

And in step seven what we did then was we spread it out. We looked at its range from minimum to maximum. We looked at the 50th percentile values, low percentiles, high percentiles, and we compared these to what we would have gotten from the so-called true maximum three-day running average. What we were looking for is how often we captured the true max and when we did not, when underestimated it, how often did that happen and to what magnitude to what extent.

Those are those questions. I can come back to that but just to show you, again this graph isn't that easy to see but what you are looking at there, each of these, look at number four. That is the bootstrapped distribution of maximum three-day averages sampled at four-day intervals.

And we went from underestimating by 31.5 percent up to getting it right on with no error. That is the range of our performance there. We want to look down at, I think, number 13 was our best run or one of the best chemographs. We went from minus two percent up to no error. So that is a very narrow range and we looked at how often. These are all again maximum running three-day averages.

And I think in some of our, let's go back up to this question. Do the highest runs equal or exceed the true max? How much lower than true are the lowest runs? And then if we wanted to create an interval on a given CWS, let me go back to this, from minus 30 to zero, that is one kind of interval. If we wanted to look down at our better ones, number 13, we go from minus two up to no error. That is a kind of interval in itself.

One of the questions that we will be asking the panel is what do we do with these as a whole? If it happens and this is totally

hypothetical, if our health effects people are looking at a three-day average and they say how well would we do if we sampled every four days, we could say, I think on mean runs, the average performance was about five percent under true. Our lowest runs, I think, were on average about 13 percent under true for sampling every four days on a three-day average.

Now we have asked them, can you live with that if you would like a four-day interval sampling strategy. The thing is we can look at this and say we average 13 percent in the lowest runs below true but we have some higher values that were 30 percent under true. So do we want to make a competence interval around our performance here or do we want to take, run them all as a whole, look at the lowest one percent of all the CWS and say 99 percent will be at or better than this kind of performance. I think that is what we were getting at in asking for advice on bounds for the population, what to do with that, or can we produce a prediction

interval because we are looking at what we have. But what do we expect on new systems?

Then when we go from here to individual samples, I think we are relying on asking your advice on things like the different interpolation levels, which did make some difference, we found, in how well we underestimated or overestimated. It wasn't a lot but I think with stair-step you can overestimate more often than not. And I believe it is possible to underestimate a little more in magnitude but we didn't do a systematic look at that.

That and any other methods that you could, kriging, whatever that would help us on the individual level. We are trying to go from talking about bounds on something like this for the population to then what do we do on an individual level.

SESSION CHAIR PORTIER: Dr. Heeringa?

CHAIR HEERINGA: Steve Heeringa. You bootstrapped this 4500 times but in a systematic

sample of 365 days, there are only 90 unique samples. So are you not simply just at random sort of getting an expected repetition of five of these?

MS. FRANKENBERRY: We repeated, yes, on purpose to get a larger number of samples so that when we looked at the lowest one percent or the highest one percent, we would have more samples to deal with but definitely the finite value for each of the running averages. And we simply re-sampled more to be able to have more to deal with.

CHAIR HEERINGA: There were sort of 91 or 92 unique values that could occur from your sampling process in CWS?

MS. FRANKENBERRY: I think it is more for four days because if you start from day one to day three, then day three, two to four, three to five, all the way up to 365, 363, it is something like 365 minus three or four days, something like that. I think the worst case is with 90-day averaging. I think it is 365 minus

90. It comes around, something like that.

SESSION CHAIR PORTIER: Yes, Dr. Gilliom?

DR. GILLIOM: Again, I think this will come up later but just a comment briefly now because of the point just made. And I think this is meant as an illustration of a data analysis process, more than like the final answer. So this won't be meant to be particularly critical.

But to evaluate a short-term exposure like this, the way the truth was created is not appropriate, basically.

I am going to keep coming back. The starting point in how we define truth is the absolute most critical step in every one of these experiments. Like here, truth on a daily basis was created from samples that were 30 or 35 times a year. It is missing many of the characteristics of short-term fluctuation. So when you simulate sampling from that, you are going to do pretty well recreating what you already got from a limited sampling.

So basically the underlying data used to define truth has to be denser than the type of time frame you are trying to evaluate. And that is why you see so much work done on these relatively few sights that have daily data. And we will see this as kind of a recurring --

SESSION CHAIR PORTIER: This is Ken Portier. Your underlying simulation model has to be complex enough to capture that high resolution variability.

DR. GILLIOM: Yes, you have to have some basis to defend that your truth represents truth for the timescale you are going to now experiment with.

SESSION CHAIR PORTIER: I think most people understand that.

MS. FRANKENBERRY: And we did acknowledge that, I think, in the paper. Some of the Heidelberg datasets were not quite 365 but of course they were much better than 30 samples. So and I think that was our next step in plans. We just only got so far. I think, Nelson, did you

want to address it later?

SESSION CHAIR PORTIER: It doesn't look like we have any additional questions. Mr. Thurman, did you want to have some comments?

MR. THURMAN: Yes, actually I want to bring us up to from, I am not going to say down in the BS pile but from down in the weeds, I want to bring us back a little higher elevation in terms of we get into this discussion.

And first of all, I do want to distinguish one area where we do not agree with Syngenta on this. I mean, some of the contention is the assumption is we should only be doing this assessment based on finished water because finished water through all the processes are going to be more blended and smoothed out than raw water.

However, the best data we see in finished water is weekly sampling. So we don't know what is happening in between. And I actually was pulling up an example last night for something else and it turned out to be a pretty

good example here. This is for one system in Ohio in one year. The magenta are the raw water samples. The dark blue triangles are the finished water samples. And what you see is the finished water follows the same type of pattern we see in the raw water.

There is nothing in this that suggests to me that we would expect the finished water to be less variable on the days that weren't sampled than we see in the raw water in this particular system.

So this is why we are looking at as we try to define the truth, what are the best datasets out there to define the truth. And those that have more robust sampling are what we are going to work with.

The other thing that is interesting is you look here, you see an early season shorter peak and a longer season, a larger peak later in the season. Very similar to the pattern we were showing in the Missouri site. So this is not a pattern that we would not expect to see in some

of the community water systems. And so as we are going through the analysis, we are looking at those that have more intensive sampling so that we have a better start with what the truth is.

And actually the reason I pulled that sample out in the first place is something Bob Gilliom raised and something some other panelists had asked. We have been talking about atrazine and we are looking at total chlorotriazines. The advantage of the AMP monitoring that Syngenta has done is that they have measured not just atrazine but they have measured the individual components to get at the total chlorotriazines.

I was looking at Ohio because I know there is simazine use in Ohio. So I wanted to try to show you an illustration of a site where we have got both atrazine and simazine detected. And you can see the dark blue triangles here happen to be the simazine pattern. The magenta circles are atrazine and the triangles are the total chlorotriazines.

This actually shows that simazine is

following the same pattern that we see with atrazine but I think that is because it is also being used on corn. In other areas of the country where it has different uses, we may see different timing.

We do have this data that we can test that. We can start asking ourselves do we see a different pattern with simazine than we do with atrazine and what effect would that have on there? So we do have that power.

Okay, so let's go back to where I was hoping the discussion would kind of help us, and I wanted to explain to the tox people why trying to decide are we looking at one day short-term or longer term and not knowing drives us buggy.

This chemograph shows the blue line is your daily measurement at this site. This red line shows that if we were looking at a four-day average, this is what type of -- this is your rolling four-day average concentrations. So you see, there is still a pretty good influence of the shorter day measurements in a four-day

exposure.

I started working on the 90-day exposure and I got this far and I probably didn't realize okay, I needed to go back and get more data and then decided I needed to sleep more than I needed that.

But what I wanted to show is the magnitude. This data point here and moving along starts here, starts whenever you get your first 90 days averaged in. So you can see these exposures are muted, potentially missing this is not going to have as big an impact on a 90-day exposure period as it is on a four-day exposure period. So that is why knowing what our exposure window is, really helps us in terms of how we decide and how we interpret the data. So I wanted to put that in to keep in mind, as you go along.

Now if we were to take a look at that seven-day sampling frequency that I showed it seems like weeks ago but I think it was the beginning of the week, this is the red line is

the four-day rolling average profile you see estimated with all the daily values. The blue line here is that same four-day rolling average profile estimated with the weekly sampling points that you saw in that previous slide.

So this is kind of what we are looking at. We see more of a difference, even on the shorter term averages, depending on the sampling frequency than we would otherwise. So this is the context I want you to think about as we get into these questions. And with that, I am not going to throw any more slide at you.

SESSION CHAIR PORTIER: Dr. Akana?

DR. AKANA: I have a late thought for you. In the HPA world -- Well, first of all, our small group here decided that a one-day exposure, an acute hit is probably going to be okay with atrazine. But in my personal view in the HPA world, a hit say the third cycle away, so you get a hit on the first cycle and if you get a little hit on the third cycle, we are verging into this where you get an extra effect on that third

cycle.

So now I am interested in little clusters of little spikes. So up one day, down. And then say six days later there is another one day little up and down spike. That, as I understand it, would not be picked up by most sampling of finished water because of the way the water is processed. But if that actually reached -- Well, in the lab you can do that.

It can be just as deleterious and verging into an episodic repeated exposure, which is one spike of chronic. Chronic is not just one up-peak and sustained of say atrazine. Little spikes can be bad, too.

Now I am wondering if in your work you can detect clusters of little spikes.

MR. THURMAN: Boy, that just made things --

If you sample frequent enough, yes. But that gets back to how frequent do you need to sample. And if I go back, if you are looking at something like this, even a creation -- but you

can start seeing that. If were say, and I am going to use 20 just as an example and please don't take that as this is what we are looking at, but you can start to estimate how often do we encroach that. And once again, how well we can estimate that. The frequency above a certain threshold value or a certain averaging period threshold value, we can estimate that. It depends on how frequently we have to sample to do that. So, it can be done.

SESSION CHAIR PORTIER: I wanted to make the point, too, you know, the spikes can be artificially generated by lifestyle. So suppose the background your tap water is a 20. On day one half of your two-liter is from tap water. On day two you are drinking bottled water. On day three you are drinking tap water again. You have just created that double pulse that you are looking at. And so from EPA's point of view, there is a lifestyle that has got to be integrated into this and how people get their two-liter dose every day, it is not always from

this.

Dr. Heeringa?

CHAIR HEERINGA: Steve Heeringa. For Nelson, you look at this double peak pattern which we have seen, particularly in some of the agricultural drainages. Is this typically the result of pre-emergent and post-emergent application of atrazine or is it the result of random rate of all events following application?

Because there is information. If it is pre-emergent/post-emergent, you pretty much know planning time and application times, in terms of intensive sampling. So how much do we know about that?

MR. THURMAN: I may give you a little more complex answer. I mean, it could very well be the result of when the farmer is getting out of the field in relation to the rainy period. And it is a conjecture because I don't have the rainfall data at this point. It could be that you had some initial planning going on. And so you had some initial atrazine applications for

the short -- then you get a rainy period. And so the farmer could not get back out in the field until later. And then you see the second one coming later.

It could be there is a difference in intensities of the rainfall events that you see here. This far apart suggests that you had two different application periods. And it could be for any number of reasons. A lot of times what we have seen is that it is weather-related in terms of how much can the farmer get out before the rains and when do the fields dry up enough so that they can get back out again?

SESSION CHAIR PORTIER: Dr. Krishnan?

DR. KRISHNAN: I just want to add to the discussion that the spikes and their relationship to the hits, I mean something that is in-between is the internal dose measure. And so these spikes may not necessarily translate to the spikes of the appropriate dose measure in the body that drives the sequence of events.

Given the rate of absorption and so

forth, the appropriate dose surrogate tends more like the total chloro products. So I think one of the focus would have to be, well essentially because you can relate those effects more closely to those internal dose measure, rather than to the external spikes and the internal dose measure would be more on the early end of the curve in rating these.

And one of the focus would have to be considering the integration of the drinking water input with the PBPK model so that some of the dose metric profiles can be evaluated in the context of the detail evaluation as they go forward. I think that would certainly add to the science basis of qualities evaluations.

SESSION CHAIR PORTIER: Excuse me. I have 10:17. We will take a 15 minute break and then we will see if my gamble of increasing understanding reduced the uncertainty in the discussion time.

We will return at 10:35.

(Whereupon, the foregoing proceeding

went off the record at 10:18 a.m. and resumed at 10:38 a.m.)

SESSION CHAIR PORTIER: Mr. Thurman, I guess you are reading the questions. Hey, we are not on question 1.9 anymore. Okay, Dr. O'Byrne?

DR. O'BYRNE: Could I just ask one very brief question? I was very surprised at your graph that you plucked out last night from wherever that the level of atrazine in the finished drinking water showed the same profile as raw, if I understood it properly. And you sort of used that as evidence to criticize Syngenta for focusing on finished water. I may be misinterpreting this.

I am absolutely amazed that you see the same profile because it depends on where that finished water came from. I mean, was it a big pool? A small pool? Because I would thought there would have been a massive dilution of anything.

MR. THURMAN: And I am not sure I

would call it a criticism. I think it is a distinction between the way we look at things and the way we interpret things. We have some smaller community water systems that don't have much of a holding period where you do see the atrazine moving through and unless you have sufficient carbon filtration, the atrazine will continue to move through the system.

Some of the smaller systems I think you heard Alan Roberson talk about that in the public comments, they do try to treat. And some of the systems we do see where the treatment is knocking down the atrazine levels, sometimes because carbon filtration is expensive, they don't necessarily have it all year round. They try to time it as best as they can. Sometimes they get it, sometimes they miss.

Sometimes you see in some of the systems where for the most part it is down but then you will a spike coming through. But there are a few systems that we are looking at that do have a similar profile to what we see.

And the point we are making is we don't necessarily want to wholesale right off what we think are some valuable robust datasets because they are not finished water sets. It is what Bob Gilliom is talking about, how you defined the truth. And if we have the robust data sets that have sampling patterns and shapes and patterns that are similar to what we are seeing in these community water systems, we think those are still very valuable. In fact, more valuable to develop our statistical analysis and the approach we take to evaluating the monitoring strategies because they do capture that frequency that is smoothed out when you have weekly sampling or less frequent sampling.

So that was the point I wanted to try to make in terms of that distinction. It is not meant as a criticism but I think the reason we are not writing off well water samples just because they are not finished drinking water.

SESSION CHAIR PORTIER: This is Ken Portier. It probably also reflects their

conservative nature. Right? So they look for worst-case scenarios and they do their risk assessment from a worst-case scenario, figuring that that is going to be protective for everybody else who are in better case scenarios from a public health point of view.

Question 2.1.

MR. THURMAN: Okay. In conjunction with the toxicological review presented in the issue paper, the Agency has also discussed methods for re-evaluating the sampling frequency that is necessary for determining, with confidence, concentrations of the pesticide in water that sources drinking water. These have included different methods for estimating pesticide concentrations between known sampling events and examining the performance of different sampling strategies for averaging periods of different durations. The Agency seeks feedback from the Panel with regard to how the uncertainty and variability in both the monitoring data and in the toxicity data (i.e., the point of

departure) can be integrated to characterize and to interpret the potential significance of atrazine concentrations in drinking water.

Given the nature of the temporal patterns of pesticide occurrence in surface waters described in Section 5.2 of the issue paper, including serial correlations from day to day, periodicity in elevated concentrations within seasons and from year to year, detections below quantitation data, and uncertainty in the shape of the pesticide distributions in surface waters, what statistical approaches should the Agency consider in determining confidence bounds on exposure estimates from monitoring data? Please comment on how the approach may vary depending on the duration of concern.

SESSION CHAIR PORTIER: Dr. Young?

DR. YOUNG: Well, considering confidence bounds on exposure estimates for monitoring data, a key consideration is what is being estimated. Now that sounds pretty simple but here that seems to be quite a challenge.

So what you do differs on whether we want to estimate a specific quantile given daily data, or rolling average data, or if we want to estimate the peak or whatever we want to do.

So, I want to put that out there first because what we suggest will differ depending upon what the final decision is.

So, if say I want to -- spikes seem to be important. And if I want to estimate a particular quantile, say the 99.99 quantile, the 99th percentile over the course of a period of time, and let's suppose that is my goal for right now. Okay? And if I want to say -- I think another question that has to come up with this particular atrazine, with this particular application, is whether you want that for the whole year or for a concentrated period of time in which the concern is greatest. In other words from prior to the start of planting to harvest or sometime in that time frame. Because you know that in most years, it is not as great of a concern.

The easiest way to do such a sample set in those confidence bounds is to know the distribution. And then you take advantage of the properties of the distribution to set confidence intervals.

The problem is I don't think anyone here is comfortable with knowing what that distribution of values is. And when you do not know that, then you are pretty much moving toward non-parametric approaches. And I see that both in that EPA does and Syngenta does and I think that is the right approach.

Now, in order to actually use non-parametric approaches, one of the big things that comes up is sample size. And basically, you have to have enough data if you want to estimate those extreme quantiles with any precision.

So if you want a 0.95 quantile, you need at least 20 observations, at least. The standard there may be unacceptably large for that sample size but that is a minimum and if you want a 0.99, you need a hundred. Okay? That is just

the reality of what you have to have in order to do a decent job.

And I saw a table of the sample sizes in one of the Syngenta documents somewhere here. And I put a similar one but I think -- and also it is in the ILSI report that you referenced in the Appendix A. I thought the Appendix A in that report did a very nice job of outlining options and certainly those could provide great guidance.

In that report, some simulation was done and found that based on the same thing we are seeing is interpolating among values and dense datasets. And what they found was that there was a tendency to overestimate the quantiles, the extreme quantiles. So that is a little conservative. And I believe the reason that is the case is that inherent in these methods is the assumption of a random sample and the presence or correlation results in the effective sample size being smaller than that taken. And so we are actually pushing it out a bit more.

And so I think work, that is a hard problem, but I think work on how to actually figure out what the effective sample size is and refine those methods may be something you are interested in but it is somewhat comforting from the public standpoint that it is a conservative estimate at the present time.

If we move to these rolling averages, then the problem is further complicated because normally we think in an independent sample, the variance goes down. You just divide by the sample size as far as the mean goes but you have this positive correlation, which tends to keep that variance inflated a bit.

And so that would need to be considered in these methods. I have also thought a little bit about using an extreme value theory in this setting and especially on the rolling averages because as soon as you begin averaging, then you can begin to appeal to perhaps the central limit theorem and if the data aren't too skewed, then maybe you can begin to use some of

those ideas to reduce sample size. I haven't fully explored that but I think it is something worth looking into.

SESSION CHAIR PORTIER: Thank you. Dr. Coupe is next.

DR. COUPE: Thank you. I think first off I need to clear up something. This corner of the table has been referred to as hydrologists or hydrology. There is actually only two hydrologists and two statisticians.

(Laughter.)

SESSION CHAIR PORTIER: That is an important point, yes.

DR. COUPE: Yes, I spent an hour being told what the difference between Bayesian and a frequentist is. I still have no idea.

So my remarks are going to be, I am not going to touch too much on the statistics but mostly on the observations on the hydrology of what we are talking about.

To begin so there is an issue on what data should be used to determine exposure

assessments. There is data on atrazine concentrations from the intake, the stream or reservoir side, and then there is finished water data. the U.S. EPA has decided to use the intake values for assessment of human exposure.

The gentlemen from CropLife and Syngenta suggested that the more appropriate data to use to evaluate human exposure is the finished water values. I can see both sides of this argument as there can be considerable difference between intake concentrations and the concentrations in the finished water. However, there are a number of studies showing that a treatment plant effectiveness in removing atrazine is variable depending upon many factors, which includes the type of water, the organic matter content or the pH, the type of treatment, how well trained the treatment plant operators are, the maintenance of the treatment. In some cases atrazine survives the treatment process relatively unaffected. Given this, I think it is appropriate to use the intake values to estimate

the exposure to humans with the understanding that this is a conservative estimate and that the actual exposure may be less.

Given that, I do think Syngenta made a valid point in bringing up the fact that the community water system, that data that we are shown by the U.S. EPA, would not have been used for drinking water, as that plant selectively pumps from the stream into a holding pond and from there into another larger holding pond.

And I think this model is more common in the small community water systems that take directly from a stream. They don't have intakes on the reservoir. And in this case, the appropriate place to sample the water used by the plant to evaluate human exposure, would have been from the intake from the larger pond. The data as shown gives a misleading impression of high exposure from this community water system.

There are uses for these data, of course, such as ecological exposure but it seems inappropriate for the question being asked here.

The duration of concern makes a big difference to the sampling strategy. If the duration of the concern is very short, that intensive sampling during expected high concentration period is probably the only real answer. But as the duration of concern increases, the sampling intervals can be eased and if the duration gets long enough, you can actually estimate it from models such as WARP.

And in general, it seems that most sites in the AMP program run by Syngenta, the variability and the amplitude of the data are not that great. If you look at page 94 where it states in the White Paper from the U.S. EPA, it states that 90 to 96 percent of the data are less than three parts per billion. And Syngenta shows pretty much the same thing in their response on pages seven and eight.

Now this is good news in the sense, especially when you consider these data are biased with more frequent sampling during expected periods of high concentration. In fact,

the AMP data suggests that there is only a subset of some 27 or so of highly vulnerable community water systems.

And to my mind, these sites probably need to be treated separately and more carefully. For the other sites, the 120 or so, the concentrations are relatively low and don't exceed the MCL often. But I suspect that is because of the storage dynamics of these reservoirs and these community water systems probably have low --

Okay, what I wanted to say was that these community water systems probably have consistently low levels of atrazine in their source water throughout the year because of storage dynamics. And realistically, probably more people are exposed to low concentrations of atrazine for longer periods of time than they are higher concentrations at shorter periods in time.

I thought of one other things when we were talking, I think Dr. Akana brought it up and a couple of other people were talking about the

spikes, and so far mostly what we have seen in our chemographs are, I suspect just very small sites. And so they show a very sharp one or two peaks.

If you move to larger sites, you can have multiple peaks. Especially if you take a look at the Missouri data, you can see multiple peaks during drinking water season kind of depending on whether it is raining in Iowa or it is raining in Missouri, or raining in Nebraska. So you can have, in one community system, you can have a number of small spikes but they can consistently come through your system.

Thank you.

SESSION CHAIR PORTIER: Thank you. Dr. Gilliom?

DR. GILLIOM: I'll just supplement, rather than repeat some of what has been covered. The first part of the question is related to what approaches should be used to consider in determining confidence bounds on exposure for monitoring data. And so far we have talked

mostly about the use of site-specific data to estimate those confidence bounds. And I just want to make the point that there is also a range of more indirect approaches using inference from existing data. And there is basically probably two categories of that. One you have heard a lot about in relation to the sampling experiments of taking these highly sampled data sets and driving relationships that show how confidently we can estimate a particular value with a given sampling frequency.

So those are kind of categorical things. The Crawford 2004 paper and the ILSI appendix are good examples of that, which have then been extended further by Syngenta and EPA and they have plans to do even more.

So those give you a good initial estimate in many cases of what kind of confidence bounds to expect on any particular concentration statistic. And you can re-do it for other ones if you need to so it gives you an idea of how much you have to worry about it.

The second category which we have also just had mentioned here is the WARP model type approach, which is really just a multiple regression approach to relate basing characteristics to concentrations and create predictions. That is another indirect way to make, based on the data we have already accumulated from many sites, make estimates of both the concentration statistic and the confidence bounds for unmonitored sites.

And I bring those type of methods up mainly, and particularly the second one on the regression models is that a big part of the approach here in the end is going to be to identify the relatively small proportion of sites that need the most intense energy applied to them for the high sampling problems and so forth.

And this gives us a direct and quantitative way of getting at that, both what the expected concentration statistic will be and what the expected confidence bounds on that will be, which can now be done.

And then the second part of the question is related to the duration of concern and how the approach might vary. And I guess I just want to reiterate the point that I think has already been made a couple of times and it is probably obvious at this point but the shorter the duration of the concentration statistic, the more intense data we need to get truth.

And it seems simple but it just reiterates how much the answer that we come up with on monitoring design is driven by how you define very specifically the concentration objective. So the Agency in the end is the one who has to do that from what you all say but we need to know the amount, the time frame, and the confidence bounds required to meet the requirements of the Agency and then you can design around that.

SESSION CHAIR PORTIER: This is Ken Portier. Just to clarify, when you say intense data, you mean temporally dense measurement.

DR. GILLIOM: I will say in a more

general way, temporally intense in relation to the sought concentration objective.

SESSION CHAIR PORTIER: Dr. Lee?

DR. LEE: I don't have too much to add to what has already been said. I just want to make two small points. One is that if we are talking about confidence bands, I think it is important to take into account some of the sources of variability that we are not looking at right now like measurement error.

You know, if you go and measure the stream, if two different people do it, you are going to get somewhat different answers probably. And that sort of variability needs to be taken into account, if you want to make a precise confidence statement. And that doesn't necessarily have to be estimated on a stream-by-stream basis. I think we can learn a lot about that sort of measurement globally and just apply an estimate there. But that sort of variability should definitely be taken into account.

And the second point I want to make is

that sort of follow on to some of what has already been said, that the way you are going to estimate the confidence band really depends on what sort of model you are using to fill in the gaps. Assuming you are not doing daily sampling, you need some sort of model to say how we are going to fill in the gaps and then the type of confidence band is going to follow from that.

So I do want to disagree somewhat with Syngenta's statement that we don't need modeling. It is a model of some sort, even if we are doing linear interpolation. That is a model. And so either we need to sample daily or we need to use some sort of model to fill in in-between. And most likely, it is not a need to sample daily because we can fill in in-between. But it requires some model and that is going to affect how we are going to compute confidence bounds.

SESSION CHAIR PORTIER: Thank you. We will open it up to the rest of the panel. Comments? Dr. Hayton?

DR. HAYTON: Do we have any idea of

how big a daily dose is the threshold of concern? Because I am showing these chemographs showing a hundred parts per billion. That is a hundred micrograms and in two liters, 200 microgram daily dose. Would that ring any bells in the BPA/BPH signaling network? And to me, if it doesn't, do we need to catch those peaks?

SESSION CHAIR PORTIER: Dr. O'Bryne?

DR. O'BYRNE: I don't think we have any evidence at all that the small amounts have any effect. They are orders of magnitude greater.

SESSION CHAIR PORTIER: And I think that has been said a couple of times before. Maybe that is a subheading to our final report.

Any additional comments? Yes, Dr. Heeringa.

CHAIR HEERINGA: Steve Heeringa. The question may be for EPA. Thinking about the simulations that you have done and that Syngenta did and we get this sort of rough projection from a simulation done on some periodic sampling to at

least truth as defined in an empirical data set and we get factors like 1.5 or 1.24. How would, in the context of where you set these limits, we have uncertainty factors, is that sort of uncertainty built into what you would think of as your typical uncertainty factors or is this something additional?

MR. THURMAN: Okay, we have wrestled with how you do with uncertainty. I think you have heard, for instance, that atrazine you have heard the value is 12.5 and 37.5 used. When we are looking at the weekly monitoring of these community water systems, we have been looking at how is that, because it is more intensive sampling, we have been comparing that to a 37.5, with less intensive sampling because there is uncertainty in that we have been looking at the 12.5. So there is kind of a 3X that has been used in that regard.

So it is possible that we might hold that monitoring uncertainty, if you will, in as part of the overall uncertainty or safety factors

that we look at.

DR. LOWIT: Anna Lowit. Just to add to that, to sort of talk about the source of where that comes from, as atrazine is regulated under the food quality protection act, the FQPA has a provision that requires the application of the 10X. And that 10X accounts for both hazard and exposure and that 10X value can be reduced, as Nelson was talking about, the three, based on information that the Agency looks like and the science supports doing that. It can also be increased.

So I just wanted to make sure that was explicit of where that would be derived from.

SESSION CHAIR PORTIER: Dr. Reed?

MR. THURMAN: And by the way, this is why you need hazard and exposure folks working together on this.

DR. REED: I just want to put in a caveat about the comparison, the water level and then two liters per day and compared to the animal study, there is a lot of "uncertainties"

because you have the interspecies variability considered and also inter-individual consideration. And so a straight comparison probably is not very productive.

SESSION CHAIR PORTIER: Any additional comments? Okay, I think we have gotten some good comments, Dr. Young and the four. Dr. Heeringa?

CHAIR HEERINGA: Just one additional comment. You know, I support what Dr. Young indicated, too, and I think others that Dr. Gilliom with regard to sort of differential sampling over the years. I mean, if we think about it simply as a sampling problem from a frequentist's perspective, we would essentially allocate sample to these intervals in proportion to the standard deviation, the measures within the intervals. And this can be adaptive, too. In other words, you could start with a periodicity and depending on the water system and the information that is gleaned over time. So I think rather than trying to hit a home run the first time, try to get into the ballpark and then

think about refinement as you collect additional data.

SESSION CHAIR PORTIER: And this kind of was a question that came up as I was reading this. This is Ken Portier.

When these community water systems take these samples and process them, what is the lag time between the sample and the number actually being received back to the process, the CWS manager? Is that hours, days or months?

DR. COUPE: All of the above.

SESSION CHAIR PORTIER: Really? What is the median on that? Is that weeks?

(Laughter.)

SESSION CHAIR PORTIER: I didn't want to max and min. I am more of a median. Is that like weeks or days?

DR. COUPE: Oh, it is probably on the order of days but there are some very large systems that have large holding times and there are very small systems that --

SESSION CHAIR PORTIER: Shift it off.

Excuse me. What was that? Weeks.

Because you know, -- yes, ma'am? Please identify yourself.

MS. BISCOE: Melanie Biscoe. I am the CRM or Chemical Review Manager for atrazine.

I think Syngenta has been talking amongst themselves a little bit back here, and we actually discussed this earlier this year. So about ten days in the EPA, Syngenta monitoring, CWS monitoring program, that you actually get the numbers back.

SESSION CHAIR PORTIER: So this has an impact for some of the things that statisticians and samplers would like to say which is well, you could do adaptive sampling. Once you start to see it go up, you increase the sampling rate. But if there is a ten-day lag and many of these windows are 20 to 30 days wide, that methodology is probably not feasible here. And then you are stuck with kind of choosing a window of time and increasing your sampling during that window of expected peaks and decreasing it.

So the current methodology is feasible within the practical limits of the sampling and turn around time.

Now of course if you had a dip stick methodology and a color code that said high, you could use something like that to increase your sampling, even to within a day. I mean, you know, if it really was moving fast, you could be picking samples hourly.

Dr. Heeringa?

CHAIR HEERINGA: Steve Heeringa. Just to be clear, when I was talking about being adaptive, I was thinking the feedback might be annual. But as usual, Dr. Portier has a shorter cycle on these things than I do.

SESSION CHAIR PORTIER: So you are talking about using last year's profile to tell -- The unfortunate problem with that, of course, is climate.

So you know, a lot of this does seem to have a relationship to rainfall. And so while there is a persistence in climate, it is not

strong enough that you can really use that.

CHAIR HEERINGA: But I would say that there are multiple variables in this work model, including how much crops planted, the absorption of the system, flux through the system. That probably has more temporal permanence than the climate. We are not going to predict the rainfall but we can predict roughly when people are going to plant crops, when the atrazine is going to go on the field and roughly how systems are going to respond in the cross different rainfall events.

So I think we are averaging over and we don't want to over adapt. But I am just saying that if you were in a system that clearly showed much more variability than the sort of averages on which you are basing your sampling plans, then there might be an argument over time to intensify sampling for that system, until you were confident that you were getting what you wanted for that system.

SESSION CHAIR PORTIER: This is Ken

Portier. And I think we are only repeating recommendations we made at the SAP a year ago when we looked at the ecological issues related to environmental sampling. So you can go read those.

Dr. Gilliom.

DR. GILLIOM: Just, one comment here since it is not really brought up in the charge questions later, and it is on this issue of whether there is some short-term adaptive approaches that could be taken. And I don't think it is worth getting heavily into this at this point.

But if it was to turn out that there is a really short term concentration objective that is extreme, like a one or two-day type of thing, then it really would be possible to use a quick screening process, like an amino assay test for atrazine triazines and just do it in the water plant and make decisions.

So there are tools. They can trigger a laboratory analysis and they are well-known and

well-characterized. So if that becomes a big issue, I think it is a discussion item that should be thought through.

SESSION CHAIR PORTIER: Make sure you add that to the report.

Okay, I think we have stretched this one out as long as we can. Why don't we move to question 2.2?

MR. THURMAN: I would be glad to. The first two simulation methods presented in Section 5.5 are applicable to the specific data sets they describe, although some generalities regarding shape patterns appear to exist. Given this information, please comment on the strengths and weaknesses of the approaches and on the practical merits of pursuing them or some other numerical approach with a larger set of higher concentration systems. Please comment on how the methods for determining confidence bounds might apply given these considerations.

SESSION CHAIR PORTIER: Dr. Gilliom.

DR. GILLIOM: Okay so some of this has

been talked about a lot. So I am going to kind of abstract from my comments.

Both the methods that are referred to in here follow the same concept we have talked about this morning. And generally it is the concept described in the Crawford paper 2004 in which the actual data for selected sites are interpolated between samples, and then treated as truth, and then sampling experiments are done from them.

So I think as I said before, probably my comment would sum up that the approach makes sense if the initial actual data are sufficient to simulate reality for the problem at hand.

And so in this specific example, just to pick on the two ones that are in here as examples, is that they are not, they are examples of simulated truth from real data that are not adequate to address a short-term concentration objective. So these were examples generated from 30 to 40 samples a year and then they were interpolated to create a simulated truth, and

then they were subsampled to test how they did for doing shorter term concentrations. It is like a seven-day or four-day average or whatever.

So and I think everybody realizes that but they are actually good examples of what to watch out for. The analytical process is fine but the specific application is not appropriate for short-term.

And so I think the general point I would leave it with is that yes, the basic approach is good. We have to make sure that once the concentration objectives are defined very precisely and probably more precisely than anybody is going to feel comfortable with on the biology end but the Agency is still going to have to do it is that then we design a sampling analysis process that fits that.

And I think I will just leave that there.

SESSION CHAIR PORTIER: Dr. Coupe?

DR. COUPE: I don't think I have too much to offer to Dr. Gilliom. I just say there

isn't a great method for extrapolating these data. You have what you have and if you need to find a resolution then you just need to collect more, collect measurements.

SESSION CHAIR PORTIER: Dr. Lee?

DR. LEE: I don't think I really have anything to add to this one. I will have more to say probably later about modeling and bringing it in but I will hold that to number four.

SESSION CHAIR PORTIER: Dr. Young?

DR. YOUNG: This is one that probably from your view, unfortunately, I feel strongly about.

The real question that is not answered that really has to be answered before we know how to simulate is how important are the peaks and how long do the peaks last? And we don't really have that information right now. But there is one thing that I think needs to be made really clear and I think we have hit around it several times but it so important I want to hit it again.

If one wants to draw inference at the

daily level, based solely on the data collected, sampling must occur at least daily. If one wants four-day rolling averages, you can't have a rolling average without a least two values. So the minimum would be two in those four days and that may not be enough.

Simulations, models or any other approach that suggests otherwise is making some strong assumptions about what happens at the finer timescale.

Now, if a good understanding of the systems exists, then one may be able to model the results, in which case sampling could be confirmatory. But this requires knowledge of the system and sufficient supporting data to construct such a model for each side, something that is not present here, at least not yet.

It does seem to be reasonable to concentrate sampling effort during the time of exposure. However, as Bob has noted several times, the sampling scheme developed for atrazine may not be applicable to other contaminants.

The thing that bothers me about all the simulations that we have seen here is that they are smooth versions of reality and that is true even for the daily data that we have. There is no accounting for measurement error, variability within the day, or all the other things that happen when you are out in the field. And anyone who has been there knows how bad that can be.

It would seem to me that a more reasonable approach for simulations is to try to bring some modeling information such as the WARP to generate chemographs with typical behavior and then to bring the variability associated with that modeling process to bare. So you would have a realization of the chemograph followed by an analysis in that realizations would be replicated so you would capture all of the variability in the system. You wouldn't just be hinging on one particular realization of the chemograph. Because even if it is real, it is dead and gone.

That is not about what is happening in

the future and I think we need to capture that variability in the simulations because I think as a consequence of doing it the way we have been doing it, it makes everything look better than it should and that is a little scary.

So that is, I think, the primary point I want to make.

SESSION CHAIR PORTIER: Additional comments?

This is Ken Portier. When I looked at this, you know, to me the strength of the methodology as Linda mentioned in the first question is that is its non-parametric nature. So a lot of it you are going kind of back to first principles of sampling. You are simulating. So to me, that is the strength of the process.

The weakness is the starting data. It is clear. You know, if the starting data is an inadequate representation of what is really happening that the water system, nothing that we can do statistically is going to improve that

starting data with any kind of first principle statistical methodology. You are going to have to go to modeling or more sampling. I mean, it is kind of that simple.

Now, I personally like the WARP approach because I know that there is other data out there, climate data, soil data, that is sampled even more regularly than these water system data is sampled. And so through modeling we can use the correlation relationship and be able to impute a little more of what is going on.

And so I do like that kind of approach because it just uses more data, uses the information. And that includes kind of modeling from the well water to the water that comes out of the community water system. I haven't seen any modeling talking about the effectiveness or the methodology.

I mean, you talk about activated carbon systems but do those fail when the concentrations reach a certain level. I mean, you know, some of these systems work well when

there is just a little bit coming in. But then if the ability of the carbon to capture these molecules gets overwhelmed by the concentration, you could actually have a discontinuous peak kind of event happening as well. So I think you need that kind of. And that is probably a whole other division in EPA that you have to go and talk to. Right?

Mr. Thurman?

MR. THURMAN: I do want to point out on the drinking water treatment, we have actually came to the SAP a number of years ago on drinking water treatment and what we knew, what the state of the published literature was on that. We have continued to keep an eye out on that and update that to the extent we do.

Generally when we are doing drinking water exposure assessments that would be used in dietary exposures, we addressed drinking water treatment effects as separate at another end of the process. So we didn't bring this to you because we addressed them separately but it is a

point well taken.

In fact, I looked at it, and you are starting to look at sources of variability. You throw in the drinking water treatments, you start to expand your sources of variability that you have to address. So we were trying to take a simpler approach then we can layer that on. But that is a point we will.

SESSION CHAIR PORTIER: I mean, I recognize as a public health risk assessment methodology, you can put limits on raw water because as long as you can argue through that raw water back to a consumption and a risk.

So I mean I recognize that as long as you have kind of got that continuum explained, we can still go back and say well but we don't want to see any input, raw waters above a certain level because we know that gets translated for all these processes, with all these uncertainties into levels that are not safe. So, you know, it is just that, and you guys know.

Any additional comments? Dr. Gilliom.

DR. GILLIOM: Just to add to the thought on the model application. I think we are in a position to very effectively use the available models built from existing data to identify with known probability the sites that need the attention for a specific objective.

So there is a lot of points in there to follow but once the objective is defined we can, in other words, estimate which specific sites. And then I think we have to remember in evaluating all these methods, that and it is the agency's direction, I think appropriately, that once you are to that point, each individual system on its own becomes the focus, not some aggregate statistics for that group or anything. It is that system and those people that use that system.

And at that point, the problem can be further evaluated as needed, including adaptive sampling or in some cases as brought up in examples, differences between the intake screen water, holding reservoirs and finished water.

Because at this point, you are down to a very small proportion of the community water supplies.

So it is a part of an overall decision-making process that right now we are only talking about he FIFRA add-on, that is what I call it anyway, to regular compliance monitoring. We haven't even, I mean, a compliance monitoring is not involved yet. It is still over there on quarterly sampling per year but it could eventually be affected by this down the road if the Office of Water chooses to do that.

SESSION CHAIR PORTIER: Okay, I don't see any additional questions. I think you got a pretty good answer on that one.

MR. THURMAN: I think we got an answer.

SESSION CHAIR PORTIER: Let's move to 2.3.

MR. THURMAN: As described in Section 5.4.2 of the issue paper, the Agency is considering the use of a confidence interval or

prediction interval approach to characterize the uncertainty of exposure estimates derived from monitoring data of varying sampling frequencies. Please comment on the strengths and weaknesses of either placing confidence bounds on the rolling average estimates and comparing to the upper limit from monitoring against the level of concern or, conversely, placing confidence bounds on the LOC.

And I apologize. That was written by committee. I hope you will be glad to clarify that.

SESSION CHAIR PORTIER: Dr. Coupe, you are first up. It looks like there were four people assigned to these four questions and then Joe used the random number generator to assign who was first.

DR. COUPE: When I saw the list of questions come out without names attached, I looked at 2.3 and said, "Dear God, I hope I don't get that one."

(Laughter.)

DR. COUPE: I think I understand the second part of this. And I will paraphrase and you correct me if I am wrong but you are talking about in the second part of taking the uncertainty in creating an LOC and then creating a bound around that and then testing your data against that.

MR. THURMAN: That is correct.

DR. COUPE: So, I just leave the particulars of the statistics details to my colleagues doing a 95 percent confidence interval slapped over another 95 percent confidence interval. But I don't really think you want to do that, just given the variability of how an LOC is derived and the different safety factors and moving from animals to humans. I think you probably wanted to stick with the LOC number.

But I am going to move into a little more broader thing. I am talking about the variability of using a statistical method to determine future sampling scenarios. I know we had to do that. But it assumes that the

distribution, the constituent of instrument is going to be the same from year to year and we know it is really not true.

There are many things that change from year to year that might change, that change the distribution of atrazine in the surface water of our basin. So when we talk about year to year variation in rainfall as well as long-term climate change, the changes in crop types, for example, you get a new cold weather variety that can be planted earlier or later.

We have a lot money going into the Mississippi River Basin now for BMPs and these have the ability to change how the atrazine is moved into the surface water. And you are going to have changes in the weed population and the weed infestation which have changed your herbicide use.

But I just wanted to give you a little brief recap of herbicides in water, just kind of a 101 on the distribution of atrazine in surface water. The distribution of atrazine in surface

water can really be explained in two terms, a source strength and a hydrology. A source strength being the fact that for atrazine to be detected in the surface water of a basin, it has to be used in that basin. We have mentioned work before but if you look in WARP, the single largest parameter that explains variability and explains more variability than all the rest of the parameters combined was use. So, was atrazine used in the basin? If it was, then some small proportion of the applied amount ended up in surface water.

And the other important factor in the transport of atrazine is water. There must be water to move atrazine into the stream of the basin. For the most part, atrazine is also transported atmospherically but the concentrations here are moved off the landscape by water.

So the EPA has stated how hard it is to relate concentrations to the streams of flow and I agree that it is because the concentration

is made extremely high with a very small runoff event soon after application and there may be a very small concentration, a large runoff event a few weeks later. But I submit that hydrology can explain the presence or absence of atrazine in streams.

Additionally in larger streams, the variability of concentration may be due to the timing of the arrival of water at the intakes from various parts of the basin. Some may be having more rainfall than others. Or maybe planning was further along in one area of the basin than it was in another one.

There is an awful lot of information and expertise on the transport of agricultural chemicals and I am convinced, as opposed to determining if, atrazine poses a health risk. It is not rocket science. A program can be designed that will do what is needed to do if we know what is needed. What you need to know and have to know to design a program is what is being discussed here as what is the endpoint. At what

level does the atrazine concentration in water matter to humans or the environment? It makes a big difference if it is 25 ppb or 0.25 ppb.

That being said, here is a few more random comments. I don't think a one size fits all approach is necessary. Crawford showed in his paper that the larger the basins, the less data you need to reach precision goals. Conversely, of course, the smaller the basin, the flashier it is and the more samples are needed to ensure precision. Given that there is significant history at these sites and many have shown only small levels of atrazine, I believe the sampling strategy could be tailored for individual community water systems, probably less samples on the larger systems and maybe more on the smaller ones.

And earlier, there was a discussion about daily sampling. I just want to iterate that as my colleague Linda put it. If you need fine resolution, and she said it several times already, then you have to sample at a fine

resolution. If you want a four-day average, I thought it was a very good point, a four-day average, you have to sample twice. Yes, you can't really create, interpolate or model data if that is what you need. I whole heartedly agree with that and that goes for even sub-daily sampling, which we haven't even mentioned about because some of these basins may have variability that relies on the sub-daily.

So and also just one last point as we mentioned before. I would like to point out that although there is a lot of very good work on the toxicity of atrazine, from my own experience and as the woman from NRDC stated, you never find atrazine alone in a water sample. There is usually a plethora of other constituents, some herbicides, other kinds of materials that are in there. You just never find atrazine alone.

So to study the toxicity of atrazine alone kind of short-sheets safety. I think it is a serious concern. Thank you.

SESSION CHAIR PORTIER: Dr. Gilliom?

DR. GILLIOM: I guess on the response to that specific question, you know, my tendency is to not try to put additional, this is an agreement with Dr. Coupe, try to put confidence bounds on the LOC. It seems by its nature it is a process of putting in safety factors and coming up with a conservative threshold. So at least once that is all done, you do have a fixed value to compare something to. I don't think we should start trying to do even more with that.

I do think it is really important that we have agreed upon and predictable ways that confidence bounds are going to be put on the exposure estimates. And part of the reason I say that is that I think in the end we are going to have a whole continuum of approaches being used to estimate exposure. So you could envision potentially a high probability group of sites that you have very dense actual measurements on and very tight estimates of confidence on one hand. And those are going to have a different probability of exceedance issue than if we are

using an indirect method from a model, which is also going to be an important part of the continuum.

So I would kind of view it in terms of we would like to see, I would like to see the objectives stated in terms of acceptable probability of exceeding a given threshold. That way, we can look at any method on a common playing field. So in other words, if the EPA gives us guidance and says okay, the moving average value is seven, seven-day value of 20, and it is okay with us that there is a 50 percent chance that that is exceeded, then we can look at any estimation method on a common basis and evaluate the probability on those terms. And that is not easy to do but in the end, that is what it kind of comes down to, even if you have a very shallow basis for estimating the numbers.

SESSION CHAIR PORTIER: Dr. Lee?

DR. LEE: Yes, I concur with my colleagues here. And I just want to say that statistically the problem of placing confidence

bounds on a level of concern is a harder problem than just putting confidence bounds on a rolling average. And there is going to be enough complication perhaps in other things that we want to do with this, that this may not be the place to add additional complication.

SESSION CHAIR PORTIER: Dr. Young?

DR. YOUNG: I have nothing to add.

SESSION CHAIR PORTIER: That just increased my probability of getting through this morning. Anyone else?

Yes, I tended to concur also. I mean I thought about this and I thought to myself exactly as you say, it is probably a lot harder to put a confidence interval in a limit of concern and it leads to more public confusion. The exposure estimate, everybody assumes exposure is going to be variable so a probability statement on exposure is probably much more acceptable than one on a public health limit of concern. So I think you have got a pretty clear answer to this question. It kind of gives you

the direction.

MR. THURMAN: Yes, we appreciate that. Okay, so question 2.4. We just scared Don off.

SESSION CHAIR PORTIER: He had warned us that he had to leave for another meeting.

MR. THURMAN: I am just picking on him. I will pay for that later.

SESSION CHAIR PORTIER: We won't see him again. Right?

MR. THURMAN: Well he has got six months to forget about that.

Okay, this is, it looks like a two-part question. I will read them both and you can -- three part. Oh, gosh. Okay. I apologize in advance for that.

Please comment on the relative merits of the various modeling approaches the Agency described in Sections 5.4.1 and 5.6 for interpolating pesticide concentration between sampling points and, in particular, on the strengths and weaknesses of these methods as the frequency of samples decreases.

Considering the health endpoints being considered for atrazine, particularly data for the HPA axis, and the exposure time frame needed to induce the health effects, which is shorter than that used in the 2003 risk assessment, please comment on the advantages and disadvantages of each model for evaluating the likely occurrence and exposure via drinking water of short, moderate, and long durations.

Please comment on the Agency's proposed approach for evaluating these methods, as described in Section 5.7.1. To what extent should the Agency consider other factors, such as the shape of the chemograph(Section 5.5.3), weather patterns, stream flow, and/or pesticide use patterns in evaluating the modeling approaches?

SESSION CHAIR PORTIER: Dr. Lee?

DR. LEE: You gave me the long one. All right. Let's charge in.

Section 5.4.1 describes two basic methods for filling in values between the actual

measurements one-year interpolation, stair-step imputation. Neither of these methods can ever give you a predicted value that is larger than any of the observed values. So this clearly will lead to underestimation of the maximum value, if the maximum value does not occur on a sampling day. And this carries over then into shorter term averages or any average as well. The shorter the average, the more important this peak is.

The stair-step method has, I think, further danger of missing the truth here. In terms of following the curves, you are trying to get an average, the stair-step method will tend to overestimate a decrease in curve when it is concave -- will tend to overestimate a decrease in curve. Linear interpolation also will tend to overestimate if the decrease is concave but by not as much as the stair-step method.

From the examples of chemograms that were given in Figure 7, I get the impression that most curves generally will be concave from more

of the year then convex because there is an initial peak following the application of the pesticide, followed by proportional decrease from the peak. And maybe my hydrologist colleagues can correct me on that but that is the impression I get of these shapes.

So among these two methods, the linear interpolation does seem like it is probably going to do a little bit better. But for the most part, it probably doesn't matter. We are missing the maximum significantly. There is the potential for missing the maximum significantly here.

For longer term averages, like a 90 day average or 26 week average, both the linear interpolation stair-step methods seem to work reasonably well because the underestimation of the peak values can be balanced by overestimation of post-peak values. It is not exactly a ringing endorsement but central limit theorems kicking in for us there.

But to accurately estimate a maximum

value when it is out of sample, it is going to require use of a method that can predict a larger value than those that are observed in the data. And example of such method is an artificial neural network as described in 5.6 and Appendix C.

Let me just briefly mention again what I brought up on, I guess it was Monday. Appendix C does describe the importance of not using too many modes because over-fitting is not good for prediction. Absolutely correct. But if you are using too few nodes, then you will also potentially be not fitting the curve very well. The shape of the peak may not be correctly categorized and you may be missing the maximum value as was shown in the difference between the three-node fits and the four-node fits in the White Paper.

So the importance of finding the right number of nodes is critical for getting good estimates and some sort of basic model selection like a BIC measure would be helpful. That may

also be able to eliminate the need for fitting autoregressive errors. I think the current approach that involves the neural network with autoregressive errors is just going to be too complicated for a non-expert to implement and thus, it is not necessarily a practical approach but perhaps if we can eliminate the need for autoregressive errors, it may become a more useful approach in this context.

Section 5.4.1 also mentions a number of other potential approaches: bootstrapping, kriging, random function models, regression-based models and deterministic models. Bootstrapping methods again are never going to be able to predict a value that is larger than you actually observed. So, that leads to definite worries. The other four methods do have promise alone or particularly in combination.

Kriging, which is the basic case of fitting a Gaussian process model interpolates the data with a smooth curve but does allow the curve to move outside the bounds of the data. And so

it could be used for estimating a maximum that occurs outside of the sample base.

Also like a Gaussian process, like a neural network can be used to smooth noisy data, rather than doing strict interpolation. That is an issue that I don't think has been discussed here. The Agency uses interpolation a lot in the White Paper but there is a difference between strict interpolation which necessarily will go through all the observed points and some sort of curve fitting, which may discount exact values because of say measurement error. And so the curve will get close to the points but it will do some smoothing. The neural network approach is an example of smoothing approach. It is not guaranteed to go through all the data points and we wouldn't necessarily want it to go through all the data points. But the interpolation methods, the linear interpolation stair-step are, by definition, interpolation. They are guaranteed to go through the data points. Kriging, in its basic form is an interpolation method. It will

go through all the data points but generalizations to Gaussian processes allow more flexibility if that is not necessarily the case.

Let's see. I did actually get some examples of the community water system data from Marry Frankenberry and was able to fit some basic kriging models to them and found that well, it doesn't actually do a whole lot better than linear interpolation in most cases. In most cases, I was not able to get a fit that gave me maximum values outside the sampling days that were higher than the observed ones.

And my guess just from really basic, you know, I didn't have a whole lot of time to do the analysis in the last two days, my guess is that the shape of the peaks are very sharp. And as such, the correlation structure is somewhat different around the maximum than it is in the rest of the space. And so fitting a standard stationary model as kriging would do, is not adequately characterizing the curve. A more full analysis would involve a non-stationary

correlation model and these do exist. But again, I think this is going to be far more complicated than you would want to implement on an individual community water system out in the field.

One last issue around kriging is that you do need to estimate a correlation structure. These are difficult to estimate. Empirical chemograms are highly variable and so I would recommend the estimation be done more globally, pooling across years and across water systems. Okay, that is kriging.

Random function models are another approach we can use. Essentially there we are picking a shape for the curve. And we do know a fair amount about what these curves may look like, although they differ from system to system. Using those shapes can really aid in the determination of a maximum value and can help them say in determining if the maximum value may have occurred on a sampling day or a non-sampling day. And if it is a non-sampling day, be able to estimate how much higher is that peak.

And there are a number of ways to do this. There is an example in the context of pesticide concentrations given by Vecchia et al. in a 2008 paper, essentially Dr. Gilliom is on that one, essentially using the WARP model to look at predicting -- it is combining the work model with seasonal shaped functions to be able to make predictions about where that peak might be.

So that ties into regression-based models. And there has been a number of work by Dr. Gilliom and others at looking at regressions to predict maxima and quantiles and there is a lot of potential there. They do, however, are looking at sort of the yearly total. So just looking at the maxima over the whole year, it is not going to give you a time series on its own there. And the accuracy may not be quite at the level that one would want for an individual water system but I think there is a lot of potential there. In particular, there is a lot of potential for combining these regression-based

models with other sorts of models to get improved information and help us make out-of-sample predictions.

The final method is used, the final method mentioned in the White Paper is deterministic models. These are built from a combination knowledge of the physical and chemical laws of the process, looking at the actual physical process. And those can be really useful for predicting maxima and short-term averages when we don't observe the data directly.

But an important issue there is calibration, which involves the setting of inputs and possible tuning parameters so that the predictions do closely match observed values. And there is the concern that this may need to be done for each watershed individually and then that becomes a complicated problem again.

I guess one global theme here is you don't get something for nothing. If you want to be able to make predictions in-between the observed days that are more accurate, it takes

more work.

I do want to mention one other category of models that has not been mentioned in the White Paper and that is using extreme value theory. There is a fair amount of theory in the statistics literature about modeling extreme events and their distributions developed around those. And I think Dr. Young has mentioned some of these earlier. But there are some, and there is some very recent work. There is a paper that just came out in the Journal of the American Statistical Association that looks at methods for modeling extreme values of a correlated process. Most of the literature involves independent samples but some of the more recent work does look at correlated processes such as chemograph. And so that could be really useful to be looking at. It is a very new literature and I am not that familiar with it but I think the EPA should at least investigate that literature.

MR. THURMAN: If you could put that in your report --

DR. LEE: I will put all the references --

MR. THURMAN: -- that would be great.

DR. LEE: -- in the minutes. Absolutely.

Okay, so I think that gets through parts one and two of this question. Part three then is about the procedure for evaluating the effectiveness of the different methods given in Section 5.7.1. And the general approach I think is sound but as mentioned before, it is important to make sure that we are using a truth that is realistic and so has enough level of detail and variability that reflects what we will actually see.

And then the important point is the last part of the question asks about other factors. And I think those really, really should be looked at, taking into account possible covariates, like weather patterns, stream flow, pesticide use patterns, would really help in being able to make better estimation or also the

shape of the chemograph.

So combining these sorts of other pieces of information would really help. Trying to estimate the maximum, just looking at the data non-parametrically, we can do that but it is not as powerful as bringing in other information that we do have available.

And again, Dr. Gilliom has been involved in some work that relates those pieces of information and I think that would be a critical direction for the agency to further investigate.

SESSION CHAIR PORTIER: Thank you. Dr. Coupe?

DR. COUPE: I think anything I wanted to say I have already said. And that was pretty comprehensive. So that is it for now.

SESSION CHAIR PORTIER: Dr. Gilliom?

DR. GILLIOM: Just I guess one thing to make clear is between Dr. Lee and myself, we will make sure all the references are in there for the various articles. They will be included

in the write-ups.

One, I guess point I want to make is that if the, and it is probably just a repeated implication of the discussion here but if the duration is on the very short end, it is going to be evident pretty quick, I think that we don't have enough range of conditions covered of existing examples of intensively monitored sites. So we have a few sites from Heidelberg College. We have got a couple of drinking water sites. It is growing but we don't really have the full geographic range of conditions represented. So there will be some important decisions to make there about what that means about how far we use inference from existing data. But there is no point in addressing that or even trying, I think, until we know the concentration objective from the toxicology side from the Agency.

And I think that is enough to add for now. Thank you.

SESSION CHAIR PORTIER: Dr. Young?

DR. YOUNG: I also thought Dr. Lee did

a nice job in summarizing. One of the things that kind of pull some ideas that have already been stated that maybe pulling them together is that it seems to me it makes really good sense to use the WARP or some other model to identify the most vulnerable community water systems.

And then once you found those, then it is probably worth some time and effort to figure out exactly what should be done for those systems. And the methods that have been proposed, interpolation methods, might work fine as long as a 90-day rolling average is fine. But if we begin shortening them up, they are not going to be sufficient.

So some way to get a more realistic chemograph is important and that would seem to call for maybe the simplest I can think of is a regression-type model where you put in some kriging and then once you have that, you can use geostatistical simulation to get an idea of the true variability associated with that and maybe begin putting some bounds. And then you have a

fairly good idea of what might happen within that system.

Now that would take more time but if we narrow the scope of the systems, then perhaps you have more time for individual efforts. And then once you get -- the first one always takes the longest. So that is something to think about.

SESSION CHAIR PORTIER: Any additional comments from the panel?

One of the things I thought about, we haven't really talked too much about the rolling average methodology. But you are using a fairly simple approach of just taking a couple of points and doing the regular average. And there is a slight improvement you can do on that, which would be more of a weighted average. I was thinking Linda, isn't it more like a lasso-type approach that still gives you a good average? But you notice on the graphs that Mr. Thurman showed that with the rolling averages, their average profile is always going to be shifted to

the right of the real profile and that is a function of just doing a simple average, rather than having a slightly wider window with some weights, some decreasing weights on either side.

So there is kind of a weighted smoothing that will give you kind of a similar rolling average but one that kind of matches up in terms of its peaks and its valleys with more of the original profile.

The other thing I was thinking of, a comment Dr. Gilliom mentioned about, more intensive sampling from more sites, is that it would be nice to have some "normal sites." The interquartile range sites. You know, to sort of ensure the public that we haven't just looked at worst-case scenarios. We have also looked at some good players, some solid citizens in the middle. CWS's that don't have all these major problems that are overall managed, that our methodology works well in those kind of solid citizen sites as well. You always aren't always in the extreme because you get charged with being

too extreme at that point. Right? And not really giving us a good picture.

And as far as the White Paper, you know, when you looked at the neuro network and showed us how well the smooth to the neuro network worked and then you added in the autoregressive two component, you didn't show really how well that improved that process. And the point that Dr. Lee was making on some other models is that kind of adding is nice statistically and it may improve the R-square three percent but when you looked at the picture of the smooth, it probably isn't something that is really noticeable. And so if you are looking to simplify, statistically that is nice but in terms of complexity, it really adds a lot of complexity to the estimating process to be able to do that. That may not pay off in the long-run.

I think that is the end of my comments. Dr. Heeringa.

CHAIR HEERINGA: Just one minor

statistical observation, too, which I think people probably recognize but we didn't mention and that is in the sampling process itself, regardless of the periodicity with which you draw single samples, the sort of every nth day sampling, systematic sampling would be most efficient if you had sort of long-term monotonic trends, increases or decreases. But if you have arbitrary fluctuations on shorter terms, that systematic sampling may actually give you greater variance then something that is more randomly perturbed within the fixed windows.

And I think that between what RTI did and Syngenta and what Dr. Sielken did and what you have done in your simulations, you might actually be able to see that one some chemographs. I don't know if it has been structured that way but you could actually test that.

But it is just a small point but I think systematic sampling is great if you are on a monotonic trend but could in any given sample

lead to greater variance or error than a randomly perturbed sort of fixed window sample. It is just a minor issue but is probably something technically not to lose track of in the process.

SESSION CHAIR PORTIER: Yes, Dr. Gilliom?

DR. GILLIOM: Just a tangent that reminded me of, that it is maybe important to how we translate ultimately the monitoring requirements from the toxicological requirements is just to remember that all of the data we are looking at so far and what is normally done are instantaneous grab samples for the most part. There is a few data sets that have auto samplers that are doing composites either flow or time waved and so forth.

But if it is an important thing to capture for instance to know that we have a time-weighted daily value or a time-weighted two-day or three-day value, that is important information to have in the LOC. Because too often, that type of information is left out and then the

monitoring design goes ahead with variants based on instantaneous samples and all that and it may not capture what you really want it to for the biological effects. So I think we can deal with all that, the Agency can deal with all that but we need to know the specific objective from the biological point of view.

SESSION CHAIR PORTIER: I am sitting here thinking we need a synthetic drinker that drinks two liters of water out of it in a day and then composites that and gets an average concentration. That's okay.

Any additional comments?

So that, I think is the last question.

MR. THURMAN: Mercifully so.

SESSION CHAIR PORTIER: And before I close, we usually do two things, and one is I am going to open it up to the panel for any final comments from any panel member. If there is some topic you felt that hasn't been brought up that you wish to comment on, we can add this in at this point. I think we pretty intensively

covered a lot of these topics but we will open it up to anyone.

Last chance to say anything before I turn it over to EPA for their closing remarks. I don't see anybody dying to present a new issue. Dr. Lowit, I know she has a few closing comments.

DR. LOWIT: Before I speak on behalf of the team to give you our appreciation and our sort of next, what we will be doing now, I will speak on behalf of myself.

A little story. A number of years ago, probably eight maybe nine at this point, I had only been with the Agency a year or so, two years at the most and it was the night before the first big meeting I ever gave a big presentation. A group of us were meeting with the office director at the time. It was several office directors ago. This person was going around the table giving a little pep talk and it came to my turn for a little pep talk. And the comment that I was given was you will do great. Just don't be flip. And I will forever hold that.

And yesterday I think I said a couple of things that were flip, unfortunately. And I always think of that day because it absolutely nailed my shortcomings. But I did, I think, make a comment that could have, around the science issues that was probably interpreted by all 20 of you 20 different ways and I just want to take a second and clarify something that I said.

At some point yesterday as we were talking about the new review. I don't remember if it was in the point of departure or in sort of the 101 thing that Nelson and I did, I made a statement to the effect that in this analysis we were starting from scratch or with a new slate or something to that effect. Let me just clarify what I intended because there was some context there.

Back in the fall our AA, assistant administrator, announced that the Agency, the Pesticides Office, would be doing this special reevaluation of atrazine. As part of that, we would do two major things. The first one is

focusing heavily on 2003 forward, which is the hundred and plus study that you have all been through and the reconsideration of the drinking water monitoring. The other thing that we committed to do was to ensure that the old risk assessment was safe.

There are a lot of ways to interpret that. You can interpret that that you go through every single millions and millions of pages that have been submitted to the Agency and that have been performed by researchers all over the world or other extreme is that you take the overviews and you just pick them up and you go. And I would think that some of you probably had that interpretation or somewhere in-between depending on your personal perspective.

I just wanted to make sure that this is what we are doing and it is relevant because of some comments that Penny had commented on yesterday and we wanted to make sure that you understood what we were doing. We have a small army in pesticides terms. They are putting a lot

of resources to this and I have the honor of keeping the train on the tracks. Our small army has a very large task in a very short time frame. And so what we are doing is doing what the AA asked of us. We are going to go through the old data in what we consider to be sufficient to ensure that the points of departure and the uncertainty factors in the new risk assessment are safe for human health for every life stage we can find across different durations.

That does not mean that we are going to go through millions and millions and millions of pages. What it does mean, however, is that when we select our points of departure, when we come back in September, that we will have been through enough of those pages and pages that we feel confident that we sit here to say that our new proposal is safe. And our new proposal covers all sensitive groups.

What that means in practice we have four months to figure out but I can tell you we will start with what we call the data evaluation

reviews. Yes, data evaluation records, what we call DERs which are essentially summaries of the studies that come from the guidelines. And the guideline studies are submitted, literally come in volumes. A chronic bioassay can have easily six, seven, eight, volumes. So we are talking mountains of paper.

And so what we do is take those mountains of paper and summarize them. And fairly lengthy reviews there can be easily -- some of them are 50 pages. We will start with those DERs and go from there. But we will however, focus a great deal of attention on the new studies that are coming in part because we know ahead of time that the doses are lower, the endpoints are precursor events, and so we are pushing the dose responses to the left on the dose response curve. And so that will be a large part of the focus but we will go back and ensure that they are safe, that we are not going to go through every individual animal and submit it to the Agency.

SESSION CHAIR PORTIER: Dr. Crisp?

DR. FENNER-CRISP: I wasn't wishing to suggest you had to go back and re-review a study from scratch.

DR. LOWIT: No and I didn't think you were. We just wanted it on the record.

DR. FENNER-CRISP: The point I was making was we may know now something a little bit more about a particular endpoint of concern, its mode of action, whatever, some nuance and/or have reached a stage where we now reinterpret certain kinds of data. And it was that mini task that I was suggesting had to be pursued in addition to looking at the new information.

Can I ask a question since I had the microphone? It reminded me in this last thing about the exposure time frame needed to induce is shorter than in the 2003, that refers to the chronic number, I presume.

DR. LOWIT: No, I refer mostly to the shorter term.

DR. FENNER-CRISP: Oh, okay. That was

what I was going to ask you. Do you have in mind now, looking at the other numbers over all of the course of exposure durations, two of the three categories were driven by data related to our discussion the last few days. The pubertal assay drove the middle number, the LH surge drove the chronic number, but these kinds of data did not drive the acute number.

So, could we interpret the possibility that you may be looking at this body of data related to the MOA and its consequences for generating an acute number?

DR. LOWIT: I don't know. I am not going to answer that.

DR. FENNER-CRISP: Okay.

DR. LOWIT: I am sorry, Penny, but I am going to answer a different question. I am being flip again.

DR. FENNER-CRISP: I think you said that data set wasn't --

DR. LOWIT: Yes, and that is an important point and I think we did hear a fairly

strong consensus that those single 15-minute, 30-minute cort measures would not make a robust regulatory implant. I think that was one message that we got that was pretty clear.

But as I have been sitting here and I went home last night digesting what some of you have been talking, particularly Dr. Krishnan and O'Byrne trying to blend those concepts, as I sit here, I think what we really need is a different kind of assessment. I actually think we are asking the wrong questions.

And I think this is, if you follow Kannan's line of thinking around the AUC and you blend that with the chemograph idea. And then if we can do some calculations overlaying AUC and chemograph, there may be a better way of asking these questions. Instead of thinking isolated duration, think about it on an AUC basis. It is a much more sophisticated way of thinking about it and would create havoc for the risk managers.

But I think it would bring in some of the ideas that we have heard this morning about

bringing in the level of concern idea and thinking about different CWSes that seem to fit in different categories of 20 some that have a lot of hits, and a lot of them that don't have any, and some in-between, and if you get these different patterns. I think we have been thinking about it on an AUC basis. We may be able to get more of a better distribution of that way of thinking.

And so how that fits into doing an acute risk assessment in a short-term, or in the 30 days, or an intermediate term of up to 60 days or six months or whatever it is, I don't know how that fits but I think it is actually answering the question. And I don't know how to do it but we will figure it out.

DR. FENNER-CRISP: The other thing you don't have assembled yet and we talked about it today and a little bit at the end of yesterday is the human biology that is relevant to the results that have been gathered in the laboratory animal studies. And when you have that, it will better

inform the point you have just made.

SESSION CHAIR PORTIER: I think Dr. Horton and Dr. Akana will want to comment shortly. Short comments.

DR. HORTON: Okay, very short comment. I want to summarize some things made by other people that I think all come together and one is Dr. Krishnan's comment about the area under the curve plus Dr. Akana's comment that she made earlier today about the fact that one 15-minute exposure to cort may not be a significant event but multiple exposures might be. So that when you start putting all of that together, it might be a significant event so as to put these things together.

And Dr. Cooper's comment that perhaps we have been looking under the light post for the keys and I think we were driven by a mode of action based on the mammary gland tumor and the LH surge but in the result of that, a lot of our thinking and a lot of the experiments have led us to some very interesting results and perhaps when

I finally get this figure done, we can kind of move away from the lamppost and think about how to use this new information in an informed way.

SESSION CHAIR PORTIER: Dr. Akana?

DR. AKANA: Here is an idea for you. You take a small community water system. You are measuring atrazine raw and in finished water. And then in the raw sewage, you measure cortisol.

SESSION CHAIR PORTIER: She keeps dropping these gems. I don't know what to do with them.

DR. AKANA: It might go with the bootstrap method well.

(Laughter.)

SESSION CHAIR PORTIER: Dr. Lowit, you know, as I listened to all of this, you are keeping things on track. The train idea. And I keep wondering if these are two trains kind of going along together trying to get in the station at the same time or two trains coming at each other and we don't know if they are even on the same track. So they may come together or they

may pass in the night. So it will be interesting to see when we get here in September or October.

DR. LOWIT: Well the train I am driving is the one that the data is interesting. So I don't want to sit here and speculate and sort of talk free flow because that is what I was starting to do with Penny.

SESSION CHAIR PORTIER: Just thank the panel and let's go.

(Laughter.)

DR. LOWIT: Yes, I think the point is well taken that we have collectively heard an enormous amount of helpful feedback on a very diverse almost absurdly diverse array of topics. And to say that we appreciate what you have been doing and that what you do whether it is a permanent panel or first timers, it is a lot of work to come to these meetings. We know we hand you lots of pages and ask you to digest and turn around feedback on something you may have just learned about two weeks ago.

And I am always impressed when I come

to these meetings to the degree to which really smart people come together and do really amazing things and this is just another example of that. And I want to thank each and every one of you on behalf of the whole team.

We will be back in September with the next generation of what you have seen with most likely some more choices that we haven't talked about here, points of departure, that sort of thing. We will do a more explicit evaluation of life stage sensitivity that we really haven't addressed here. Peripherally we have but we will head it straight on a little bit more.

We will also be joined by your epidemiology colleagues, both on the team and some people on the panel will join us. So we expect a good conversation around thinking about how animal and human information do and don't match. And I can tell you often they don't match and that is a great challenge.

So I will thank each and every one of you. I will thank Laura, and Joe, and Charlene

and everyone else on the SAP staff. Personal thanks to my team. It is an amazing group of people.

SESSION CHAIR PORTIER: Thank you. For the panel, once the DFO makes his final comments, I want you to just leave all your papers and your computer. We are going to go next door for a short five minute process meeting where we talk about how we produce the final report.

And at this point, I want to turn it over to the DFO, Joe Bailey, for final closing remarks.

MR. BAILEY: All I want to say is to thank the panel for all their hard work, for agreeing to come to the meeting and I look forward to working with you over the next few weeks getting the report pulled together. And I thank EPA for their presentations. I think they were very well done. And I want to thank the public commenters who came to the meeting as well to present their views.

And that's it. Thank you.

Oh, yes, and we will get the report completed within 90 days from today.

SESSION CHAIR PORTIER: Thank you. That ends the meeting.

(Whereupon, at 12:14 p.m., the foregoing meeting was adjourned.)

U